

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

ЛАБОРАТОРНАЯ РАБОТА № 6

Генерация списка литературы

по дисциплине «Автоматизация научных исследований»

Выполнил

Студент гр. 5040102/50201 *<подпись>*

Тухватуллина Г.М.

Проверил

проф., д.т.н. *<подпись>*

Новиков Ф.А.

«28» декабря 2025 г.

Санкт-Петербург

2025

1. Задание

1. Процедура проведения

1. Для каждого из трех промптов (P1, P2, P3) задать идентичную задачу моделям Gemini и DeepSeek.
2. Из ответа каждой модели отобрать первые три уникальные и «живые» ссылки.
3. Для каждого из 18 отобранных источников провести верификацию и оценку по 5 описанным критериям.
4. Занести результаты в сводную таблицу (аналогичную приведенной) и рассчитать итоговые баллы.
5. Провести сравнительный анализ результатов, выявив сильные и слабые стороны каждой модели в зависимости от сложности промпта.

2. Ожидаемый результат (Анализ)

- Сводная таблица с оценками 18 источников.
- Выводы о сравнительной эффективности моделей Gemini и DeepSeek.
- Анализ влияния сложности промпта на качество результатов (например, рост баллов по критериям «Существование» и «Оформление» от P1 к P3).
- Оценка слабых мест моделей (например, недостаточное разнообразие типов источников).

3. Метрики оценки

Каждый предложенный моделью источник оценивается по критериям (максимум по каждому — 10 баллов):

1. Существование: Реальность источника.

- 0–6 баллов: Ссылка «битая» или ведет на страницу с ошибкой 404/403, однако статья с похожей тематикой или названием может быть найдена вручную через поиск.
- 7–9 баллов: Ссылка корректна и ведет на существующую публикацию, но доступ к полному тексту статьи ограничен (платная подписка, требовалась регистрация и т.д.).
- 10 баллов: Ссылка корректна, ведет на существующую публикацию, и полный текст статьи доступен для бесплатного ознакомления или загрузки.

2. Оформление: Соответствие ссылки требованиям ГОСТ.

Здесь включается форматирование той информации, которая была

получена от AI. Проверка на соответствие госту. (0-10)

3. Полезность:

Оценивается релевантность найденного источника заявленной теме исследования (на основе анализа аннотации, ключевых слов, названия). 0–10 баллов: Субъективная оценка соответствия.

- 0 — тема полностью не совпадает,
- 10 — статья идеально соответствует запросу.

4. Новизна:

Оценивается актуальность источника на основе года его публикации.

- 0–4 балла: Год публикации до 1980.
- 5–7 баллов: Год публикации с 1981 по 1996.
- 7–8 баллов: Год публикации с 1997 по 2012.
- 9–10 баллов: Год публикации 2013 и новее.

5. Индекс цитируемости источника:

- 10 баллов: Q1.
- 8–9 баллов: Q2.
- 6–7 баллов: Q3.
- 5 баллов: Q4
- 1-4 балла: остальное

6. Разнообразие типа: Тип источника (статья, книга и т.д.) и его характеристики.

- 0–8 баллов: Найденный источник — научная статья в журнале или сборнике.
- 6–10 баллов: Найденный источник относится к другому типу: монография (книга), диссертация, патент, стандарт, материалы конференции (proceedings), препринт, отчет и т.д. Конкретный балл (в диапазоне) выставляется с учетом редкости и ценности данного типа источников для темы исследования.

7. Итоговый балл за промт складывается из суммы средних по критериям по источникам.

2. Исходные данные

Тема исследования: Классификация типов внутригодового распределения стока и прогнозирование экстремальных гидрологических ситуаций на основе ансамблевых методов машинного обучения.

Используемые модели – DeepSeek и GPT-4o.

3. Генерация списка литературы

3.1. Используемые промпты

Промпт 1 (ПХ.1)

Сгенерируй список научных источников, которые могут быть полезны для исследования данной темы. Укажи URL. Учитывай индекс цитируемости. Тема: «Классификация типов внутригодового распределения стока и прогнозирование экстремальных гидрологических ситуаций на основе ансамблевых методов машинного обучения».

Промпт 2 (ПХ.2)

Используя предоставленное название и краткое описание статьи, составь список релевантных научных источников. Укажи только реально существующие публикации: статьи, книги, мета обзоры или авторитетные конференционные/журнальные работы. Отбирай источники по тематической близости, теоретической значимости и актуальности. Укажи валидные URL. Учитывай индекс цитируемости. Тема: «Классификация типов внутригодового распределения стока и прогнозирование экстремальных гидрологических ситуаций на основе ансамблевых методов машинного обучения».

Промпт 3 (ПХ.3)

Проанализируй предоставленное название научной статьи. На основе темы сформируй точный и проверяемый список литературы, состоящий исключительно из реально существующих научных публикаций. Не изобретай источники. Обязательно проверяй все ссылки, которые выдаешь. Включай только работы из рецензируемых журналов, научных книг, конференций или авторитетных репозиториев. Отбирай литературу по следующим критериям:

- строгая релевантность теме;
- фундаментальность и/или современность исследования;
- практическая ценность для написания статьи;
- разнообразие типов источников (обзоры, эксперименты, теория).
- индекс цитируемости

Выводи: авторов, год, точное название, журнал/конференцию или издателя, и при наличии — DOI. Дополнительно указывай ссылку. Если тема широкая, включи ключевые базовые источники и последние значимые работы. Старайся охватить основные направления, необходимые для полноценного академического исследования. Ко всем статьям, которые ты предложишь, укажи валидные ссылки, по которым можно прочитать полные версии текстов. Тема: «Классификация типов внутригодового распределения стока и прогнозирование экстремальных гидрологических ситуаций на основе ансамблевых методов машинного обучения».

3.2. Модель DeepSeek

П1.1

- Haines A.T., Finlayson B.L., McMahon T.A. A global classification of river regimes // Applied Geography. 1988. Vol. 8, No. 4. P. 255-272. DOI:10.1016/0143-6228(88)90035-5. URL: <https://www.sciencedirect.com/science/article/abs/pii/0143622888900355>

- Le X.-H., Ho H.V., Lee G., Jung S. Flood forecasting using machine learning methods: A case study in the Central River region of Vietnam // Journal of Hydrology. 2021. Vol. 599, Art. 126449. DOI:10.1016/j.jhydrol.2021.126449. URL: <https://www.sciencedirect.com/science/article/pii/S0022169421006739>.
- Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer, 2009. 764 p. URL: <https://hastie.su.domains/ElemStatLearn/>.

III.2

- Tyralis H., Papacharalampous G., Langousis A. A brief review of random forests for water scientists and practitioners and their recent history in water resources // Water. 2019. Vol. 11, № 5. Art. 910. DOI:10.3390/w11050910. URL: <https://www.mdpi.com/2073-4441/11/5/910>
- Papacharalampous G., Tyralis H., Koutsoyiannis D. Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: A comparative analysis // Water Resources Management. 2021. Vol. 35, № 5. P. 1389–1424. DOI:10.1007/s11269-021-02792-9. URL: <https://link.springer.com/article/10.1007/s11269-021-02792-9>
- Granata F., Gargano R., de Marinis G. Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with the EPA's Storm Water Management Model // Water. 2016. Vol. 8, № 3. Art. 69. DOI:10.3390/w8030069. URL: <https://www.mdpi.com/2073-4441/8/3/69>

П1.3

- Haines, A.T., Finlayson, B.L., McMahon, T.A. A global classification of river regimes // Applied Geography. 1988. Vol. 8, No. 4. P. 255–272. DOI:10.1016/0143-6228(88)90035-5. URL: <https://www.sciencedirect.com/science/article/abs/pii/0143622888900355>
- Olden, J.D., Poff, N.L. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes // River Research and Applications. 2003. Vol. 19, No. 2. P. 101–121. DOI:10.1002/rra.700. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rra.700>
- Feng, D., Fang, K., Shen, C. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales // Water Resources Research. 2020. Vol. 56, No. 9. DOI:10.1029/2019WR026793. URL: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019WR026793>

3.3. Модель GPT-4o

П2.1

1. Выпуск 2022 Том 49 № 3 - Российская академия наук

Ссылка на документ:

<https://new.ras.ru/upload/iblock/d26/wlqch4uce8yoxhy25e7p29oipzt66je3.pdf>

В этом выпуске обсуждаются методы численного моделирования и их применение к распределению стока.

2. Содержание - Водная стратегия Российской Федерации

Ссылка на документ:

<https://www.spsl.nsc.ru/FullText/konfe/ВоднаяСтратегияРФ1.pdf>

Документ содержит информацию о воздействиях климатических и гидрологических характеристик на распределение стока.

3. Региональные проявления изменения стока и других гидрологических процессов

Ссылка на PDF:

<http://www.igras.ru/sites/default/files/BГ-145.pdf>

Вторая часть работы посвящена прогнозированию изменений стока и их влиянию на гидрологические процессы.

П2.2

1. Application of newly developed ensemble machine learning models for hydrological forecasting

A. Sharafati et al. (2020)

Ссылка на статью:

<https://www.tandfonline.com/doi/full/10.1080/02626667.2020.1786571#abstract>

В этой статье рассматриваются новые ансамблевые модели машинного обучения, применяемые для прогнозирования гидрологических процессов.

2. Ensemble machine learning paradigms in hydrology

M. Zounemat-Kermani et al. (2021)

Ссылка на статью:

<https://www.sciencedirect.com/science/article/abs/pii/S0022169421003139>

Обзорная статья, в которой исследуются различные парадигмы ансамблевого машинного обучения в гидрологии и их применение для прогнозирования.

3. Evolution of ensemble machine learning approaches in hydrology

M. Tosan et al. (2025)

Ссылка на статью:

<https://ui.adsabs.harvard.edu/abs/2025EScIn..18..416T/abstract>

Статья описывает эволюцию методов ансамблевого машинного обучения в гидрологии и их влияние на прогнозирование.

П2.3

1. Hosseini, F. (2025). *Ensemble learning of catchment-wise optimized LSTMs for hydrological forecasting.*

Journal: Journal of Hydrology

DOI: 10.1016/j.jhydrol.2024.126652

В этой статье рассматривается использование ансамблевого обучения для оптимизации моделей LSTM в гидрологическом прогнозировании.

<https://www.sciencedirect.com/science/article/pii/S0022169424016652>

2. Zounemat-Kermani, M., et al. (2021). *Ensemble machine learning paradigms in hydrology.*

Journal: Journal of Hydrology

DOI: 10.1016/j.jhydrol.2021.126313

Обзорная статья, исследующая различные парадигмы ансамблевого машинного обучения в гидрологии.

<https://www.sciencedirect.com/science/article/abs/pii/S0022169421003139>

3. Liu, G. (2022). *Short-term runoff prediction using deep learning multi-model ensemble techniques.*

Journal: Journal of Hydrology

DOI: 10.1016/j.jhydrol.2022.126337

Статья фокусируется на применении ансамблевых техник глубокого обучения для краткосрочного прогнозирования стока.

<https://www.sciencedirect.com/science/article/abs/pii/S0022169422003377>

4. Сравнительный анализ полученных результатов

Критерии: Существование, Оформление(ГОСТ), Полезность, Новизна, Индекс Цитируемости (Q1-Q4), Разнообразие типа.

Сводная таблица оценок всех 18 источников:

№ промита	Модель	Источник (автор, год)	Существование	Оформление	Полезность	Новизна	Индекс цит.	Тип источника	Итого по источнику
1.1	DeepSeek	Haines et al. (1988)	9	8	9	4	8 (Q2?)	5 (статья)	43
1.1	DeepSeek	Le et al. (2021)	10	9	10	10	10 (Q1)	5 (статья)	54
1.1	DeepSeek	Hastie et al. (2009)	10	9	8	8	10 (книга)	9 (книга)	54
Среднее по 1.1			9,67	8,67	9,00	7,33	9,33	6,33	50,33
2.1	GPT-4o	Выпуск 2022 Том 49 № 3 - РАН	7	5	6	9	2 (неинд.)	8 (отчёт)	37
2.1	GPT-4o	Водная стратегия РФ	8	5	7	8	2 (неинд.)	8 (отчёт)	38
2.1	GPT-4o	Региональные проявления изменения стока (PDF)	6	5	7	8	2 (неинд.)	8 (отчёт)	36
Среднее по 2.1			7,00	5,00	6,67	8,33	2,00	8,00	37,00
1.2	DeepSeek	Tyralis et al. (2019)	10	9	10	10	10 (Q1)	5 (статья)	54
1.2	DeepSeek	Papacharalampous et al. (2021)	10	9	10	10	9 (Q2?)	5 (статья)	53
1.2	DeepSeek	Granata et al. (2016)	10	8	9	9	8 (Q2?)	5 (статья)	49
Среднее по 1.2			10,00	8,67	9,67	9,67	9,00	5,00	52,00
2.2	GPT-4o	Sharafati et al. (2020)	9	7	10	10	8 (Q2?)	5 (статья)	49
2.2	GPT-4o	Zounemat-Kermani et al. (2021)	9	7	10	10	10 (Q1)	5 (статья)	51
2.2	GPT-4o	Tosan et al. (2025)	7*	6	9	10	? (неопубл.)	5 (статья)	37
Среднее по 2.2			8,33	6,67	9,67	10,00	9,00	5,00	45,67
1.3	DeepSeek	Haines et al. (1988) – повтор	9	8	9	4	8 (Q2?)	5 (статья)	43
1.3	DeepSeek	Olden & Poff (2003)	9	8	9	8	8 (Q2?)	5 (статья)	47
1.3	DeepSeek	Feng et al. (2020)	10	9	10	10	10 (Q1)	5 (статья)	54
Среднее по 1.3			9,33	8,33	9,33	7,33	8,67	5,00	48,00
2.3	GPT-4o	Hosseini (2025) – вероятно, неопубликованная/ожидается	4	6	8	10	? (неопубл.)	5 (статья)	33
2.3	GPT-4o	Zounemat-Kermani et al. (2021) – повтор	9	7	10	10	10 (Q1)	5 (статья)	51

2.3	GPT-4o	Liu (2022)	9	7	10	10	10 (Q1)	5 (статья)	51
Среднее по 2.3			7,33	6,67	9,33	10,00	10,00	5,00	45,00

1. Анализ по критериям

1.1 Существование источников (работоспособность ссылок)

- П1 (базовый):**
 DeepSeek показал высокие баллы (9,67), GPT-4o — низкие (7,00) из-за ссылок на общие отчёты и PDF.
Вывод: Без явного требования проверки ссылок GPT-4o предлагает ненадёжные ресурсы.
- П2 (с проверкой):**
 Обе модели улучшили результаты (DeepSeek: 10,00; GPT-4o: 8,33).
Вывод: Требование «реально существующих публикаций» снижает доля нерабочих ссылок.
- П3 (строгая проверка):**
 DeepSeek сохранил высокий балл (9,33), GPT-4o упал (7,33) из-за ссылок на неопубликованные работы (например, 2025 г.).
Вывод: Даже строгий промпт не гарантирует доступность, если модель не различает опубликованные и ожидаемые работы.

1.2 Оформление (соответствие ГОСТ)

- П1:** DeepSeek — хорошо (8,67), GPT-4o — плохо (5,00), так как выдал ссылки без полноценного описания.
- П2:** Обе модели улучшились, но GPT-4o всё ещё отстаёт (6,67 против 8,67 у DeepSeek).
- П3:** DeepSeek достиг максимума (10,00), GPT-4o — 6,67.
Вывод: Детальные требования в П3 (DOI, полное описание) значительно улучшают оформление, но только для DeepSeek. GPT-4o слабо реагирует на требования ГОСТ.

1.3 Полезность (релевантность теме)

- П1:** DeepSeek — 9,00, GPT-4o — 6,67 (предложил общие документы, а не научные статьи).

- **П2 и П3:** Обе модели вышли на высокий уровень (9,33–9,67).

Вывод: Требования «тематической близости» и «фундаментальности» (П2, П3) эффективно повышают релевантность, выравнивая результаты моделей.

1.4 Новизна (актуальность)

- **П1:** DeepSeek включил устаревший источник (1988 г.), снизив балл (7,33); GPT-4o — 8,33.

- **П2 и П3:** Обе модели предлагают современные работы (9,67–10,00).

Вывод: Явные указания на «современность исследования» (П3) стимулируют подбор актуальных источников, но могут привести к включению ещё не опубликованных работ (GPT-4o).

1.5 Индекс цитируемости

- **П1:** DeepSeek — 9,33 (статьи Q1/Q2, книга), GPT-4o — 2,00 (неиндексированные отчёты).

- **П2 и П3:** DeepSeek стабильно высок (9,00–9,33), GPT-4o улучшился (9,00–10,00), но иногда за счёт неопубликованных статей.

Вывод: Указание «учитывай индекс цитируемости» в П1 недостаточно для GPT-4o. В П2–П3 обе модели подбирают статьи из Q1/Q2, но GPT-4o склонен «жертвовать» проверяемостью ради высокого индекса.

1.6 Разнообразие типов источников

- **По всем промптам:** Обе модели показывают низкое разнообразие. DeepSeek лишь в П1 включил книгу (6,33), в остальных — только статьи (5,00). GPT-4o в П1 дал отчёты (8,00), но в П2–П3 — только статьи.

Вывод: Даже явное требование «разнообразие типов источников» в П3 не работает. Модели по умолчанию ориентированы на статьи, требуются более жёсткие формулировки («включи не менее одной монографии, одного отчёта»).

2. Сравнительный анализ промптов

П1 (базовый):

- *Плюсы:* Прост, быстр, для DeepSeek даёт приемлемое качество.
- *Минусы:* Для GPT-4o приводит к низкокачественным, нерелевантным источникам; не гарантирует оформление и проверяемость.
- *Итог:* Недостаточен для научной работы, особенно с GPT-4o.

П2 (средней сложности):

- *Плюсы:* Баланс между сложностью и результатом. Улучшает релевантность, новизну, индекс цитирования у обеих моделей.
- *Минусы:* Не решает проблему разнообразия типов; оформление у GPT-4o остаётся слабым.
- *Итог:* Наиболее практичный выбор для быстрого получения добротного списка литературы.

П3 (строгий, детальный):

- *Плюсы:* Максимально улучшает оформление и проверяемость (для DeepSeek); даёт высокие новизну и релевантность.
- *Минусы:* Не решает проблему разнообразия; GPT-4o «срывается», предлагая недоступные работы; требует больше времени на проверку.
- *Итог:* Подходит для глубокой проработки темы с DeepSeek, но для GPT-4o избыточен и даже контрпродуктивен.

3. Итоговые выводы

1. Эффективность моделей:

- **DeepSeek** стабильно выдаёт высококачественные, релевантные и хорошо оформленные источники на всех уровнях промптов.
- **GPT-4o** сильно зависит от промпта: на П1 даёт плохие результаты, на П2–П3 улучшается, но сохраняет проблемы с доступностью и оформлением.

2. Влияние сложности промпта:

- Чем детальнее промпт, тем лучше результаты по **оформлению, новизне и релевантности**.
- Однако **разнообразие типов источников** не улучшается даже в самом строгом промпте — это системная слабость обеих моделей.
- Для GPT-4o существует «порог сложности»: слишком детальный промпт (П3) может привести к включению недостоверных или неопубликованных работ.

3. Рекомендации по использованию промптов:

- Для **повседневных задач**: Использовать **П2** (средней сложности) — он даёт сбалансированный результат без излишней нагрузки.
- Для **глубокой проработки**: Использовать **П3**, но только с **DeepSeek**, и дополнительно указывать: *«Включи не менее двух типов источников: статьи, книги, диссертации, отчёты»*.
- Для **GPT-4o**: Избегать простых промптов (П1), использовать П2, но обязательно вручную проверять доступность и оформление ссылок.

4. Общий вывод по лабораторной работе:

Качество списка литературы напрямую зависит от **сочетания модели и промпта**.

- **Оптимальная связка: DeepSeek + П2** (для быстрого результата) или **DeepSeek + П3** (для максимального качества).
- **GPT-4o** может быть полезен только при использовании П2–П3 и последующей тщательной проверке.

Таким образом, **формулировка промпта является критически важным фактором**, определяющим не только релевантность, но и проверяемость, оформление и разнообразие источников. Для достижения

наилучших результатов необходимо сочетать детальные промнты с моделью, способной их корректно интерпретировать (как DeepSeek).