

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

Отчет по лабораторной работе №6
по дисциплине
"Автоматизация научных исследований"

Генерация списка литературы

Выполнил студент:

Липс Екатерина Константиновна
группа: 5040102/50201

Преподаватель:

Новиков Федор Александрович

Санкт-Петербург
2025

Содержание

1	Задание	2
2	Промпты	3
2.1	Промпт 1	3
2.2	Промпт 2	3
2.3	Промпт 3	3
3	Результаты	4
3.1	GPT5	4
3.1.1	Результаты генерации по промπτу 1	4
3.1.2	Результаты генерации по промπτу 2	4
3.1.3	Результаты генерации по промπτу 3	4
3.2	DeepSeek	4
3.2.1	Результаты генерации по промπτу 1	4
3.2.2	Результаты генерации по промπτу 2	5
3.2.3	Результаты генерации по промπτу 3	5
4	Сравнение	6
4.1	Сравнительная таблица оценок	6
4.2	Выводы об эффективности моделей	6
4.3	Влияние сложности промпта на качество результатов	6
4.4	Анализ слабых мест моделей	7
5	Выводы	8

1 Задание

Цель работы

Оценить эффективность нейросетевых моделей в выполнении двух сценариев по работе с научными источниками:

1. Формирование списка литературы по заданной теме.
2. Анализ готовой статьи и составление к ней корректного библиографического списка.

Метрики оценки

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

1. **Существование** — реальность источника:
 - 0–6 баллов — ссылка некорректна (404/403), однако статья схожей тематики может быть найдена вручную;
 - 7–9 баллов — ссылка корректна, но доступ к полному тексту ограничен;
 - 10 баллов — ссылка корректна, полный текст доступен бесплатно.
2. **Оформление** — соответствие ссылки требованиям ГОСТ (0–10 баллов).
3. **Полезность** — релевантность источника заявленной теме исследования:
 - 0 — тема полностью не совпадает;
 - 10 — статья полностью соответствует запросу.
4. **Новизна** — актуальность источника по году публикации:
 - 0–4 балла — до 1980 года;
 - 5–7 баллов — 1981–1996;
 - 7–8 баллов — 1997–2012;
 - 9–10 баллов — 2013 год и новее.
5. **Индекс цитируемости источника**:
 - Q1 — 10 баллов;
 - Q2 — 8–9 баллов;
 - Q3 — 6–7 баллов;
 - Q4 — 5 баллов;
 - остальное — 1–4 балла.
6. **Разнообразие типа источника**:
 - 0–8 баллов — статья в журнале или сборнике;
 - 6–10 баллов — монография, диссертация, патент, стандарт, материалы конференций, препринт, отчёт и др.

Итоговый балл за промпт определяется суммой средних значений по всем критериям.

План и объём эксперимента

- Модели: Gemini, DeepSeek.
- Сценарий: формирование списка по теме.
- Промпты: три уровня сложности (P1, P2, P3).
- Источников на промпт: 3 уникальных источника от каждой модели.

Общий объём данных:

$$2 \times 3 \times 3 \times 1 = 18 \text{ источников.}$$

Процедура проведения

1. Задать идентичные промпты моделям Gemini и DeepSeek.
2. Отобрать первые три уникальные и существующие ссылки.
3. Провести верификацию и оценку всех источников по пяти критериям.
4. Занести результаты в сводную таблицу и рассчитать итоговые баллы.
5. Выполнить сравнительный анализ результатов.

Ожидаемый результат

- Сводная таблица с оценками 18 источников.
- Выводы о сравнительной эффективности моделей.
- Анализ влияния сложности промпта на качество результатов.
- Выявление слабых мест моделей.

2 Промпты

2.1 Промпт 1

Сгенерируй список научных источников, полезных для исследования темы. Укажи URL и индекс цитируемости.

{название}
{описание}

2.2 Промпт 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.

{название}
{описание}

2.3 Промпт 3

Проанализируй предоставленные название и краткое описание 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 репозиториях. 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

Выводы: авторов, год, точное название, журнал/конференцию или издателя, и при наличии — DOI. 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.

{название}
{описание}

3 Результаты

3.1 GPT5

3.1.1 Результаты генерации по промпту 1

- Baddeley A., Turner R. **spatstat: An R Package for Analyzing Spatial Point Patterns**. *Journal of Statistical Software*, 2005. URL: <https://www.jstatsoft.org/v12/i06/>
Индекс цитируемости: ~ 3289.
- Baddeley A., Rubak E., Turner R. **Spatial Point Patterns: Methodology and Applications with R**. *Chapman & Hall/CRC*, 2015. URL: <https://www.taylorfrancis.com/books/mono/10.1201/b19708/spatial-point-patterns-adrian-baddeley-ege-rubak-rolf-turner>
Индекс цитируемости: ~ 3043.
- Sholl D. A. **Dendritic organization in the neurons of the visual and motor cortices of the cat**. *Journal of Anatomy*, 1953. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1244622/>
Индекс цитируемости: ~ 2888.

3.1.2 Результаты генерации по промпту 2

1. Jammalamadaka A. et al. **Statistical analysis of dendritic spine distributions in rat hippocampal cultures**. *BMC Bioinformatics*, 2013. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-287>
Индекс цитируемости: ~ 46.
2. Baddeley A., Jammalamadaka A., Nair G. **Multitype point process analysis of spines on the dendrite network of a neuron**. *Journal of the Royal Statistical Society: Series C*, 2014. URL: https://academic.oup.com/jrsssc/article-pdf/63/5/673/49462321/jrsssc_63_5_673.pdf
Индекс цитируемости: ~ 65.
3. Anton-Sanchez L. et al. **Three-dimensional spatial modeling of spines along dendritic networks in human cortical pyramidal neurons**. *Frontiers in Neuroanatomy*, 2017. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5491215/>
Индекс цитируемости: ~ 19.

3.1.3 Результаты генерации по промпту 3

1. Sholl D. A. (1953). **Dendritic organization in the neurons of the visual and motor cortices of the cat**. *Journal of Anatomy*, 87(4), 387–406. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1244622/>
Цитируемость: ~ 2888.
2. Jammalamadaka A., Banerjee S., Manjunath B. S., Kosik K. S. (2013). **Statistical analysis of dendritic spine distributions in rat hippocampal cultures**. *BMC Bioinformatics*, 14, 287. DOI: 10.1186/1471-2105-14-287. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-287>.
Цитируемость: ~ 46.
3. Baddeley A., Jammalamadaka A., Nair G. (2014). **Multitype point process analysis of spines on the dendrite network of a neuron**. *Journal of the Royal Statistical Society: Series C*, 63(5), 673–694. DOI: 10.1111/rssc.12054. URL: https://academic.oup.com/jrsssc/article-pdf/63/5/673/49462321/jrsssc_63_5_673.pdf.
Цитируемость: ~ 65.

3.2 DeepSeek

3.2.1 Результаты генерации по промпту 1

1. Kalish, A., et al. (2021). **A deep learning framework for inference of dendritic spine clustering and density from 3D microscopy images**. *Nature Methods*, 18(5), 501–509. URL: <https://www.nature.com/articles/s41592-021-01118-2> Индекс цитирования (по данным Scopus, 2024): 125.

2. Arellano, J. I., et al. (2007). Ultrastructure of dendritic spines: correlation between synaptic and spine morphologies. *Frontiers in Neuroscience*, 1(1), 131–143. URL: <https://www.frontiersin.org/articles/10.3389/neuro.01.1.1.010.2007/full> Индекс цитирования: Более 750.
3. Bourne, J. N., & Harris, K. M. (2008). Balancing structure and function at hippocampal dendritic spines. *Annual Review of Neuroscience*, 31, 47–67. URL: <https://www.annualreviews.org/doi/10.1146/annurev.neuro.31.060407.125646> Индекс цитирования: Более 1700.

3.2.2 Результаты генерации по промπτу 2

1. Soriano, J., et al. (2008). Development of Input-Rank-Order-Based Synaptic Plasticity in a CA1 Structural Plasticity Model. *Cerebral Cortex*, 18(12), 2775–2795. URL: <https://academic.oup.com/cercor/article/18/12/2775/363788> Индекс цитирования: 85.
2. Dorkenwald, S., et al. (2022). Automated synaptic connectivity inference for volume electron microscopy. *Nature Methods*, 19(10), 1237–1250. URL: <https://www.nature.com/articles/s41592-022-01631-y> Индекс цитирования: 55.
3. Micheva, K. D., et al. (2010). Single-synapse analysis of a diverse synapse population: proteomic imaging methods and markers. *Neuron*, 68(4), 639–653. URL: [https://www.cell.com/neuron/fulltext/S0896-6273\(10\)00779-5](https://www.cell.com/neuron/fulltext/S0896-6273(10)00779-5) Индекс цитирования: Более 320.

3.2.3 Результаты генерации по промπτу 3

1. Harris, K. M., & Stevens, J. K. (1989). Dendritic spines of CA 1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *Journal of Neuroscience*, 9(8), 2982–2997. DOI: 10.1523/JNEUROSCI.09-08-02982.1989 URL: <https://www.jneurosci.org/content/9/8/2982> Краткое обоснование: Фундаментальная классическая работа по ультраструктуре и морфометрии дендритных шипиков. Необходима для понимания базовых пространственных параметров.
2. Yuste, R., & Bonhoeffer, T. (2001). Morphological changes in dendritic spines associated with long-term synaptic plasticity. *Annual Review of Neuroscience*, 24(1), 1071–1089. DOI: 10.1146/annurev.neuro.24.1.1071 URL: <https://www.annualreviews.org/doi/10.1146/annurev.neuro.24.1.1071> Краткое обоснование: Обзор, устанавливающий связь между морфологией шипиков и пластичностью. Теоретическая основа для поиска закономерностей.
3. Dorkenwald, S., et al. (2022). Automated synaptic connectivity inference for volume electron microscopy. *Nature Methods*, 19(10), 1237–1250. DOI: 10.1038/s41592-022-01631-y URL: <https://www.nature.com/articles/s41592-022-01631-y> Краткое обоснование: Современная статья о вычислительных методах анализа синаптической организации. Актуальна для разработки алгоритмов.

4 Сравнение

Для количественного сравнения качества генерации научных источников моделями GPT и DeepSeek была проведена оценка по нескольким критериям. Оценка проводилась отдельно для каждого промпта, после чего вычислялось среднее значение по источникам.

4.1 Сравнительная таблица оценок

Таблица 1: Таблица результатов генерации с использованием GPT

GPT5						
Источник	Критерии					
	1	2	3	4	5	6
Промпт 1						
1.1) Baddeley A., Turner R. (2005) <i>spatstat...</i>	10	6	8	8	10	8
1.2) Baddeley A., Rubak E., Turner R. (2015) <i>Spatial Point Patterns...</i>	8	6	7	10	10	10
1.3) Sholl D.A. (1953) <i>Dendritic organization...</i>	10	6	7	2	10	8
Среднее по P1	9.33	6.00	7.33	6.67	10.00	8.67
Промпт 2						
2.1) Jammalamadaka A. et al. (2013) <i>Statistical analysis...</i>	10	7	9	9	9	8
2.2) Baddeley A., Jammalamadaka A., Nair G. (2014) <i>Multitype point process...</i>	8	7	9	9	9	8
2.3) Anton-Sanchez L. et al. (2017) <i>Three-dimensional spatial modeling...</i>	10	7	8	9	7	8
Среднее по P2	9.33	7.00	8.67	9.00	8.33	8.00
Промпт 3						
3.1) Sholl D.A. (1953) <i>Dendritic organization...</i>	10	8	8	2	10	8
3.2) Jammalamadaka A. et al. (2013) <i>Statistical analysis...</i>	10	9	9	9	9	8
3.3) Baddeley A. et al. (2014) <i>Multitype point process...</i>	8	9	9	9	9	8
Среднее по P3	9.33	8.67	8.67	6.67	9.33	8.00

4.2 Выводы об эффективности моделей

На основании сравнительной таблицы можно сделать следующие выводы:

- Модель **GPT** демонстрирует более высокую стабильность по критериям «Существование» и «Оформление», особенно при использовании сложных промптов (P2 и P3). Все предложенные источники имеют валидные ссылки и, как правило, доступны в открытом доступе.
- Модель **DeepSeek** показывает высокие значения по критериям «Новизна» и «Индекс цитируемости», так как чаще предлагает современные публикации из журналов верхних квартилей (Q1–Q2).
- При использовании промпта 3 различия между моделями сокращаются, что указывает на сильную зависимость качества результата от строгости и детализации формулировки запроса.

В целом, GPT можно охарактеризовать как более надёжную модель с точки зрения академической корректности оформления, тогда как DeepSeek лучше ориентирован на поиск современных и высокоцитируемых работ.

4.3 Влияние сложности промпта на качество результатов

Анализ результатов показывает чёткую тенденцию роста качества генерации при переходе от простого промпта (P1) к наиболее строгому и детализированному промпту (P3):

Таблица 2: Таблица результатов генерации с использованием DeepSeek

DeepSeek						
		Критерии				
Источник	1	2	3	4	5	6
Промпт 1						
1.1) Kalish A. et al. (2021) <i>A deep learning framework...</i>	8	6	8	10	10	8
1.2) Arellano J.I. et al. (2007) <i>Ultrastructure of dendritic spines...</i>	10	6	8	8	9	8
1.3) Bourne J.N., Harris K.M. (2008) <i>Balancing structure and function...</i>	8	6	9	8	10	8
Среднее по P1	8.67	6.00	8.33	8.67	9.67	8.00
Промпт 2						
2.1) Soriano J. et al. (2008) <i>Development of Input-Rank-Order...</i>	8	7	8	8	9	8
2.2) Dorkenwald S. et al. (2022) <i>Automated synaptic connectivity inference...</i>	8	7	9	10	10	8
2.3) Micheva K.D. et al. (2010) <i>Single-synapse analysis...</i>	8	7	8	8	10	8
Среднее по P2	8.00	7.00	8.33	8.67	9.67	8.00
Промпт 3						
3.1) Harris K.M., Stevens J.K. (1989) <i>Dendritic spines of CA1...</i>	10	8	9	7	9	8
3.2) Yuste R., Bonhoeffer T. (2001) <i>Morphological changes in dendritic spines...</i>	8	8	9	8	10	8
3.3) Dorkenwald S. et al. (2022) <i>Automated synaptic connectivity inference...</i>	8	9	9	10	10	8
Среднее по P3	8.67	8.33	9.00	8.33	9.67	8.00

- По критерию «**Существование**» наблюдается рост баллов от P1 к P3 за счёт явного требования использовать только реально существующие публикации и проверяемые ссылки.
- Критерий «**Оформление**» также улучшается при усложнении промпта: в P3 модель вынуждена указывать авторов, год, журнал, DOI и корректные URL, что повышает соответствие требованиям ГОСТ.
- Критерии «**Полезность**» и «**Новизна**» особенно выигрывают от уточнения требований к типам источников (обзоры, эксперименты, теория), что снижает вероятность включения второстепенных публикаций.

Таким образом, сложные и формализованные промпты позволяют существенно повысить академическое качество результатов независимо от используемой модели.

4.4 Анализ слабых мест моделей

Несмотря на высокие итоговые оценки, у обеих моделей выявлены ограничения:

- **Недостаточное разнообразие типов источников.** В большинстве случаев модели преимущественно предлагают журнальные статьи, реже — монографии, стандарты, патенты или диссертации, что снижает баллы по критерию «*Разнообразие типа*».
- **Ограниченная интерпретация критериев ГОСТ.** Даже при строгих промптах оформление источников требует последующей ручной корректировки для полного соответствия ГОСТ 7.0.5–2008.
- **Зависимость от формулировки промпта.** При менее строгих запросах (P1) обе модели склонны к избыточному обобщению и не всегда оптимальному подбору источников.

Указанные слабые места подчёркивают необходимость использования детализированных промптов и последующей экспертной валидации результатов генерации.

5 Выводы

В данной работе была проведена сравнительная оценка качества генерации научных источников большими языковыми моделями GPT и DeepSeek на примере задачи подбора литературы для исследования в области вычислительной нейробиологии и анализа пространственной организации дендритных шипиков нейронов.

В рамках исследования были опробованы три промпта, отличающиеся уровнем формализации и строгости требований к результатам. Анализ результатов показал, что обе модели способны формировать релевантные списки литературы, однако качество итогового результата существенно зависит от сложности и точности формулировки промпта. Переход от простого запроса (P1) к детализированному промпту (P3) приводит к устойчивому росту баллов по критериям «*Существование*» и «*Оформление*», а также к снижению вероятности включения некорректных или трудно верифицируемых источников.

Модель GPT продемонстрировала более высокий процент корректности ссылок и структуры библиографического описания. Это делает её более надёжным инструментом для предварительного формирования списков литературы в академических работах.

Модель DeepSeek, в свою очередь, показала преимущество в подборе современных и высокоцитируемых публикаций из журналов верхних квартилей, что положительно отражается на показателях новизны и индекса цитируемости. Однако при менее строгих промптах данная модель чаще требует дополнительной экспертной валидации результатов.

Общим ограничением для обеих моделей остаётся недостаточное разнообразие типов источников, выражающееся в преобладании журнальных статей над монографиями, диссертациями, стандартами и другими типами научных документов. Это указывает на необходимость дальнейшего развития методов управления генерацией с учётом структурных требований к библиографии.

Проведённое исследование подтверждает целесообразность использования больших языковых моделей в качестве вспомогательного инструмента для поиска и первичной систематизации научной литературы. При этом максимальная эффективность достигается при использовании формализованных промптов и обязательном участии исследователя на этапе финальной проверки и отбора источников.