

УДК 004.832.25

URL: <https://ptsj.bmstu.ru/catalog/icec/sacip/1058.html>

## АНАЛИЗ ЛОГИЧЕСКИХ СПОСОБНОСТЕЙ ЯЗЫКОВЫХ МОДЕЛЕЙ НА ПРИМЕРЕ РЕШЕНИЯ ЗАДАЧИ ЭЙНШТЕЙНА

**У.П. Томская**

tomskayaup@student.bmstu.ru

**М.А. Захаров**

maz@bmstu.ru

*МГТУ им. Н.Э. Баумана, Москва, Российская Федерация*

Рассмотрены различные подходы к решению задачи Эйнштейна, используемые языковыми моделями, и выполнено сравнение их эффективности. Авторы анализируют результаты работы моделей, выявляя их сильные и слабые стороны в контексте логического мышления. В статье приведены примеры решения задачи Эйнштейна языковыми моделями разных семейств: Claude, Command R, Mistral, LLaMA2, LLaMA3, GPT-2, GPT-3.5. Для каждой модели проанализировано применение логических способностей при решении данной задачи. В конце подведены итоги решения задачи разными моделями и сделаны выводы и замечания о том, как каждая из моделей способна рассуждать.

**Ключевые слова:** языковые модели, искусственный интеллект, логические способности, логическое мышление, задача Эйнштейна, сравнительный анализ, эффективность

Современные языковые модели достигли впечатляющих результатов в генерации текста, переводе и ответах на вопросы. Они способны имитировать человеческий стиль письма, создавать креативный контент и даже поддерживать диалог. Однако насколько эти модели действительно «понимают» информацию и способны применять логику для решения задач? Одним из ключевых аспектов развития искусственного интеллекта (ИИ) является проверка способности к логическому мышлению и решению задач, где требуется не только знание языка, но и умение логически рассуждать. Использование логических задач, таких как модифицированная «задача Эйнштейна», представляет собой эффективный метод оценки этих качеств.

Чтобы исключить возможность поиска готовых ответов в интернете, была произведена модификация элементов задачи с сохранением ее логической структуры. Для уменьшения количества раскладов (вариантов решения) для языковых моделей, количество элементов (в данном случае складов) было уменьшено до трех.

Эксперименты с такими задачами помогают оценить, насколько хорошо языковые модели обрабатывают естественный язык и применяют логическое мышление для построения взаимосвязей и выводов на основе предоставленной информации [1, 2]. Подобные испытания не только показывают логиче-

ские возможности и ограничения ИИ в области понимания текста и логики, но и способствуют их дальнейшему улучшению и развитию.

Таким образом, анализируя ответы различных языковых моделей на модифицированную задачу Эйнштейна, можно не только оценивать их логические способности, но и открыть пути для усовершенствования механизмов их обучения. Это исследование также открывает перспективы для дальнейшего развития языковых моделей с фокусом на логическое мышление и решение задач [3, 4].

Итак, нескольким популярным языковым моделям было предложено решить следующую задачу:

*Реши логическую задачу:  
 На улице стоят три склада.  
 Программист работает в зеленом складе.  
 Инженер ест хлеб.  
 Тот, кто пьет сок, держит пуму.  
 В белом складе пьют лимонад.  
 В центральном складе едят хлеб.  
 Водитель работает на первом складе.  
 Сосед слева от того, кто пьет воду, держит медведя.  
 На складе слева от того, в котором держат тигра, пьют лимонад.  
 Водитель работает слева от красного склада.  
 В третьем складе едят пельмени.  
 Кто ест пиццу? Кто держит пуму?*

Правильное решение данной задачи представлено в таблице ниже.

Признак	1	2	3
Цвет	Белый	Красный	Зеленый
Специальность	Водитель	Инженер	Программист
Еда	Пицца	Хлеб	Пельмени
Алкоголь	Лимонад	Вода	Сок
Животное	Медведь	Тигр	Пума

Текст ответов моделей приведен в соответствие с нормами русского языка.

**Модель claude-3-opus-20240229**

**Ответ модели.** Давайте решим эту задачу шаг за шагом, используя данную информацию.

1. Водитель работает на первом складе, который находится слева от красного склада. Значит, красный склад — второй, а третий склад — зеленый (так как всего три склада).

2. Программист работает в зеленом (третьем) складе.

3. В центральном (втором) складе едят хлеб. Значит, инженер работает во втором складе, так как он ест хлеб.

4. В третьем складе едят пельмени.

5. В белом складе пьют лимонад, а слева от склада, где держат тигра, пьют лимонад. Значит, тигра держат во втором (красном) складе, а белый склад — первый.

6. Тот, кто пьет сок, разводит пуму. Значит, в первом складе пьют сок и держат пуму, а водитель держит пуму.

7. Слева от того, кто пьет воду, держат медведя. Значит, в третьем складе пьют воду, а во втором — держат медведя.

8. Получается, что в первом (белом) складе водитель пьет сок, ест пиццу и держит пуму. Во втором (красном) складе инженер ест хлеб и держит тигра и медведя. В третьем (зеленом) складе программист пьет воду и ест пельмени.

**Ответ:** Пиццу ест водитель. Пуму держит тоже водитель.

**Уточнение.** Ответ не совсем верен. Правильно ли ты рассуждаешь в пункте 6? Ведь в пункте 5 ты говоришь, что в белом складе пьют лимонад и белый склад первый?

Вы правы, я допустил ошибку в рассуждениях. Давайте исправим решение:

1. Водитель работает на первом складе, который находится слева от красного склада. Значит, красный склад — второй, а третий склад — зеленый (так как всего три склада).

2. Программист работает в зеленом (третьем) складе.

3. В центральном (втором) складе едят хлеб. Значит, инженер работает во втором складе, так как он ест хлеб.

4. В третьем складе едят пельмени.

5. В белом складе пьют лимонад, а слева от склада, где держат тигра, пьют лимонад. Значит, тигра держат во втором (красном) складе, а белый склад — первый.

6. Слева от того, кто пьет воду, держат медведя. Значит, в третьем складе пьют воду, а во втором — держат медведя.

7. Тот, кто пьет сок, держит пуму. Поскольку в первом и третьем складах уже определены напитки (лимонад и вода соответственно), то сок пьют и пуму держат во втором (красном) складе.

8. Получается, что в первом (белом) складе водитель пьет лимонад и ест пиццу. Во втором (красном) складе инженер ест хлеб, пьет сок, держит тигра, медведя и пуму. В третьем (зеленом) складе программист пьет воду и ест пельмени.

*Исправленный ответ:* Пиццу ест водитель. Пуму держит инженер.

**Уточнение.** А если учесть, что в задаче в каждом складе можно держать только одно животное, то что будет?

Вы правы, я упустил это условие. Давайте пересмотрим решение с учетом того, что в каждом складе можно держать только одно животное:

1. Водитель работает на первом складе, который находится слева от красного склада. Значит, красный склад — второй, а третий склад — зеленый (так как всего три склада).

2. Программист работает в зеленом (третьем) складе.

3. В центральном (втором) складе едят хлеб. Значит, инженер работает во втором складе, так как он ест хлеб.

4. В третьем складе едят пельмени.

5. В белом складе пьют лимонад, а слева от склада, где держат тигра, пьют лимонад. Значит, тигра держат во втором (красном) складе, а белый склад — первый.

6. Слева от того, кто пьет воду, держат медведя. Значит, в третьем складе пьют воду, а во втором — держат тигра (так как в каждом складе только одно животное).

7. Тот, кто пьет сок, держит пуму. Поскольку в первом складе пьют лимонад, а во втором и третьем уже определены животные (тигр и медведь соответственно), то пуму держат в первом (белом) складе, где также пьют сок.

8. Получается, что в первом (белом) складе водитель пьет сок и лимонад, ест пиццу и держит пуму. Во втором (красном) складе инженер ест хлеб и держит тигра. В третьем (зеленом) складе программист пьет воду, ест пельмени и держит медведя.

*Исправленный ответ:* Пиццу ест водитель. Пуму тоже держит водитель.

Из данного ответа модели следует, что в нем прослеживаются логические рассуждения, однако в моменте про того, кто пьет воду, начинает путаться. При попытке исправления забывает о том, что в одном месте можно держать только одно животное, чего в задаче напрямую не сказано.

### Модель `command-r`

**Ответ модели.** Логическая задача с кодом `colors`. На улице стоят три склада: красный, белый и зеленый. Если `گناشت` код цвета в порядке появления, то ответ будет следующим.

Зеленый склад — программист.

Белый склад — здесь пьют лимонад, а также по тексту мы видим, что это третий склад, где едят пельмени.

Красный склад — здесь работают соседи водителя, а также здесь не держат тигра, так как он на складе слева. Следовательно, красный — это центральный склад, где едят хлеб.

Следовательно, инженер ест хлеб в красном, центральном складе.

А значит, тот, кто пьет сок и разводит пуму, находится на белом складе.

**Ответ:** на белом складе держат пуму, а пиццу ест инженер из красного склада.

**Уточнение.** Ты уверен, что на красном складе не держат тигра? Подумай еще.

Да, я уверен, что на красном складе не держат тигра.

В задаче сказано, что тигр находится не на центральном складе (а красный — это центральный). Тигр находится на складе слева от центрального. А слева от красного склада — зеленый состав, где работает программист. Поэтому тигр находится не на красном, а на зеленом складе.

Таким образом, ответы на вопросы не меняются: инженер ест хлеб в красном, центральном складе; в белом складе держат пуму, а пиццу ест человек из красного склада.

Из данного ответа модели видно, что в него подмешаны слова из других языков. Логика прослеживается очень слабо, хотя есть некоторые правильные рассуждения в начале. При этом в утверждении про красный склад ошибается и не хочет исправляться даже после вопроса об этом. Оба ответа модель дает неправильно, при этом даже не отвечает на вопрос кто, а говорит, где именно.

**Модель mistral-large-2402**

**Ответ модели.** Для решения этой задачи нам нужно собрать все факты и расположить их в таблице:

Склад	Цвет	Работник	Еда	Напиток	Животное
1		Водитель			
2			Хлеб		
3			Пельмени		
	Зеленый	Программист			
	Белый			Лимонад	
	Красный				

Начнем решение:

- Из факта 7 следует, что водитель работает на первом складе.
- Из факта 8 следует, что сосед слева от того, кто пьет воду, держит медведя. Это значит, что воду пьют либо на втором, либо на третьем складе.
- Из факта 9 следует, что на складе слева от того, в котором держат тигра, пьют лимонад. Это значит, что тигра держат либо во втором, либо в третьем складе.
- Из факта 4 следует, что в белом складе пьют лимонад. Это значит, что белый склад должен быть слева от второго или третьего склада.
- Из факта 10 следует, что водитель работает слева от красного склада. Это значит, что красный склад не может быть первым.
- По пробам можно установить, что только одно распределение удовлетворяет всем условиям:

Склад	Цвет	Работник	Еда	Напиток	Животное
1	Красный	Водитель	Пицца	Вода	Медведь
2	Белый	Инженер	Хлеб	Лимонад	Тигр
3	Зеленый	Программист	Пельмени	Сок	Пума

**Ответ:** Пиццу ест водитель, пуму держит программист.

**Уточнение.** В пункте 6 ты пишешь, что красный склад не может быть первым, но в итоговой таблице он первый.

Прошу прощения, я допустил ошибку в пункте 6. Красный склад может быть первым, потому что водитель работает слева от красного

склада, а водитель работает на первом складе. В остальном решение верно.

Вот исправленная таблица:

Склад	Цвет	Работник	Еда	Напиток	Животное
1	Красный	Водитель	Пицца	Вода	Медведь
2	Белый	Инженер	Хлеб	Лимонад	Тигр
3	Зеленый	Программист	Пельмени	Сок	Пума

*Ответ:* Пиццу ест водитель, пуму держит программист.

Из данного ответа модели следует, что ответы даны правильно, хорошо прослеживается логика, ответы представлены в виде таблицы, полностью понята особенность задачи. Однако в одном из последних пунктов допущена ошибка, в результате таблица составлена неверно и при попытке исправления она не меняется. Тем не менее это не мешает правильно ответить на поставленные вопросы.

### Модель llama-2-70b-chat

**Ответ модели.** Слева направо:

1. Водитель (работает на первом складе).
  2. Программист (работает в зеленом складе).
  3. Инженер (ест хлеб).
  4. Неизвестный (пьет сок и разводит пуму).
  5. Неизвестный (пьет лимонад в белом складе).
  6. Неизвестный (ест хлеб в центральном складе).
  7. Водитель (работает слева от красного склада).
  8. Неизвестный (ест пельмени в третьем складе).
- Кто ест пиццу? — Неизвестный (ест пельмени в третьем складе).  
Кто держит пуму? — Неизвестный (пьет сок и держит пуму).

**Уточнение.** Можешь ответить на вопросы точнее?

Конечно, я могу попытаться ответить на вопросы более точно.

1. Кто ест пиццу?

*Ответ:* Неизвестный (ест пельмени в третьем складе).

2. Кто держит пуму?

*Ответ:* Неизвестный (пьет сок и разводит пуму).

3. Кто работает на первом складе?

*Ответ:* Водитель.

4. Кто работает в зеленом складе?

*Ответ:* Программист.

5. Кто ест хлеб в центральном складе?

*Ответ:* Неизвестный.

6. Кто пьет лимонад в белом складе?

*Ответ:* Неизвестный.

7. Кто работает слева от красного склада?

*Ответ:* Водитель.

8. Кто ест пельмени в третьем складе?

*Ответ:* Неизвестный.

9. Кто держит медведя?

*Ответ:* Сосед слева от того, кто пьет воду.

10. Кто пьет воду?

*Ответ:* Неизвестный.

11. Кто работает на втором складе?

*Ответ:* Неизвестный.

12. Кто работает на третьем складе?

*Ответ:* Неизвестный.

13. Кто ест хлеб на втором складе?

*Ответ:* Неизвестный.

14. Кто ест пельмени на третьем складе?

*Ответ:* Неизвестный.

15. Кто пьет сок?

*Ответ:* Неизвестный (пьет сок и держит пуму).

Из данного ответа модели следует, что логика в рассуждениях прослеживается крайне плохо, ответы даны крайне неопределенно и даже после второго шанса не было попытки выстроить логическую цепочку — было просто написано «неизвестный».

### Модель llama-3-70b-instruct

*Ответ модели.* Давайте разбираться по шагам.

1. Программист работает в зеленом складе.

2. Водитель работает на первом складе и слева от красного склада, значит, первый склад — не красный. Значит, первый склад — зеленый, а программист — водитель.

3. В центральном складе едят хлеб. инженер ест хлеб, значит, инженер работает в центральном складе.

4. В белом складе пьют лимонад. На складе слева от того, в котором держат тигра, пьют лимонад, значит, белый склад слева от склада, где держат тигра.

5. Тот, кто пьет сок, держит пуму. Тот, кто пьет воду, имеет соседа слева, который держит медведя. Значит, тот, кто пьет сок, не пьет воду, иначе у него был бы сосед слева с медведем.

6. В третьем складе едят пельмени. Первый склад — зеленый, второй — центральный, где едят хлеб, значит, третий склад — белый, где пьют лимонад.

7. Теперь можно расставить по местам:

- 1-й склад (зеленый): водитель-программист;
- 2-й склад (центральный): инженер, едят хлеб;
- 3-й склад (белый): пьют лимонад, едят пельмени.

8. Осталось найти, кто ест пиццу и кто держит пуму. Поскольку для всех остальных видов еды и напитков есть соответствующие склады, значит, пиццу едят на том складе, где держат тигра (иначе на нем бы ели пельмени или хлеб). А тот, кто держит пуму, пьет сок и работает на складе, соседнем справа от склада, где держат медведя.

Таким образом, мы получаем:

- 1-й склад (зеленый): водитель-программист;
- 2-й склад (центральный): инженер, едят хлеб;
- 3-й склад (белый): пьют лимонад, едят пельмени;
- склад, где держат тигра: едят пиццу;
- склад, где держат медведя: пьют воду;
- склад, где держат пуму: пьют сок.

Кто ест пиццу? — Те, кто работает на складе, где держат тигра.

Кто держит пуму? — Те, кто пьют сок.

**Уточнение.** Нужно дать ответ, кто именно по профессии ест пиццу и кто держит пуму. Также задача предполагает, что программист и водитель — это разные люди и они живут в разных местах.

Извините за недопонимание. Давайте продолжим решение задачи:

1. Программист работает в зеленом складе.  
2. Водитель работает на первом складе и слева от красного склада, значит, первый склад — не красный. Значит, второй склад — красный.

3. В центральном складе едят хлеб. Инженер ест хлеб, значит, инженер работает в центральном складе.

4. В белом складе пьют лимонад. На складе слева от того, в котором держат тигра, пьют лимонад, значит, белый склад слева от склада, где держат тигра.

5. Тот, кто пьет сок, разводит пуму. Тот, кто пьет воду, имеет соседа слева, который держит медведя. Значит, тот, кто пьет сок, не пьет воду, иначе у него был бы сосед слева с медведем.

6. В третьем складе едят пельмени.

7. Теперь можно расставить по местам:

– 1-й склад: водитель;

– 2-й склад (красный): ? ;

– 3-й склад (белый): пьют лимонад, едят пельмени;

– 4-й склад (зеленый): программист.

8. Осталось найти, кто ест пиццу и кто держит пуму. Поскольку для всех остальных видов еды и напитков есть соответствующие склады, значит, пиццу едят на том складе, где держат тигра (иначе на нем бы ели пельмени или хлеб). А тот, кто держит пуму, пьет сок и работает на складе, соседнем справа от склада, где держат медведя.

9. Теперь можно расставить по местам:

– 1-й склад: водитель;

– 2-й склад (красный): держат медведя, пьют воду;

– 3-й склад (белый): пьют лимонад, едят пельмени;

– 4-й склад (зеленый): программист;

– склад, где держат тигра: едят пиццу, инженер;

– склад, где держат пуму: пьют сок, ...?

Таким образом, мы получаем:

Кто ест пиццу? — инженер.

Кто держит пуму? — программист.

Из данного ответа модели следует, что в целом логика прослеживается довольно хорошо. Изначально не совсем понята особенность задачи (не было учтено, что программист не может быть одновременно водителем), но после исправления модель приходит к почти правильному ответу (дан правильный ответ, что пуму держит программист), хоть и забывает об условии, что складов всего три и придумывает четвертый склад.

### Модель gpt2-chatbot

**Ответ модели.** Для решения задачи сначала установим порядок складов и распределим известные факты по местам. Обозначим склады как 1, 2, и 3, где 1 — первый слева, 3 — последний справа.

1. Водитель работает на первом складе.

Склад 1: водитель.

2. В центральном складе едят хлеб.

Склад 2: хлеб.

3. Инженер ест хлеб.

Склад 2: инженер, хлеб.

4. В белом складе пьют лимонад.

Склад?: лимонад, Белый.

5. В третьем складе едят пельмени.

Склад 3: Пельмени.

6. Водитель работает слева от красного склада.

Склад 1: водитель (не красный).

Склад 2: красный.

7. Программист работает в зеленом складе.

Склад?: программист, зеленый.

8. На складе слева от того, в котором держат тигра, пьют лимонад.

Склад?: лимонад.

Склад?: тигр.

9. Сосед слева от того, кто пьет воду, держит медведя.

Склад?: вода.

Склад?: медведь.

10. Тот, кто пьет сок, держит пуму.

Склад?: сок, пума.

Теперь расставим оставшиеся элементы:

— Водитель на первом, значит красный склад это второй. Так как в центральном складе (втором) едят хлеб и это красный склад, значит, белый склад — третий, где пьют лимонад.

— Значит, в первом складе пьют воду (сосед слева от водителя держит медведя, значит, в красном складе — медведь).

— Следовательно, в складе слева от того, где держат тигра, пьют лимонад, это белый склад, значит, в красном складе держат тигра.

— Программист в зеленом складе, значит, первый склад зеленый.

— Тот, кто пьет сок и держит пуму, должен быть в третьем складе, так как первый и второй уже заняты водой и медведем, а также лимонадом и тигром соответственно.

Итак, распределение следующее:

Склад 1: водитель, зеленый, вода, медведь.

Склад 2: инженер, красный, хлеб, тигр.

Склад 3: программист, белый, сок, пельмени, пума.

*Ответы:*

— Кто ест пиццу? На основе данных пиццу никто не ест.

— Кто держит пуму? Программист держит пуму.

**Уточнение.** А если учесть, что пиццу ест тот, кто не ест ни пельмени, ни хлеб?

Учитывая дополнительное условие, что пиццу ест тот, кто не ест ни пельмени, ни хлеб, мы можем пересмотреть распределение еды и профессий. Ранее мы определили следующее:

- склад 1: водитель, зеленый, вода, медведь;
- склад 2: инженер, красный, хлеб, тигр;
- склад 3: программист, белый, сок, пельмени, пума.

Так как на складе 2 уже едят хлеб, а на складе 3 — пельмени, то единственным оставшимся вариантом для пиццы является склад 1:

- склад 1: водитель, зеленый, вода, пиццу, медведь.

Теперь мы можем точно ответить на вопросы:

- Кто ест пиццу? Водитель ест пиццу.
- Кто держит пуму? Программист держит пуму.

Таким образом, учитывая все условия и исправления, мы имеем полную картину распределения профессий, напитков, еды, и животных по складам.

Из данного ответа следует, что модель очень хорошо справилась с поставленной задачей. Было небольшое недопонимание особенности задачи, но после небольшого разъяснения ответы на оба вопроса были даны верно. Однако если оценивать полную картину распределения цветов и напитков, то она не совсем верная.

Также была предпринята попытка задать данной модели исходную задачу Эйнштейна, с которой она тоже справилась очень хорошо, ответив правильно на все вопросы с первого раза и поняв особенность задачи без дополнительных разъяснений. Это может говорить о том, что данная модель была обучена на данной задаче.

### Модель gpt-3.5-turbo-0125

**Ответ модели.** Из условий задачи следует:

1. Программист работает в зеленом складе.
2. Инженер ест хлеб.
3. Тот, кто пьет сок, держит пуму.
4. В белом складе пьют лимонад.
5. В центральном складе едят хлеб.
6. Водитель работает на первом складе.
7. Сосед слева от того, кто пьет воду, держит медведя.
8. На складе слева от того, в котором держат тигра, пьют лимонад.

9. Водитель работает слева от красного склада.

10. В третьем складе едят пельмени.

Из этих условий можно сделать следующие выводы:

— Водитель работает на первом складе, значит, первый склад красный.

— Водитель работает слева от красного склада, значит, водитель работает на втором складе, который зеленый.

— Следовательно, центральный склад белый.

— По условию, в центральном складе едят хлеб, следовательно, инженер ест хлеб на центральном складе.

— Сосед слева от пьющего воду держит медведя, значит, на втором складе, где пьют воду, держат медведя.

— В центральном складе едят хлеб, значит, водитель, который работает на втором складе, ест хлеб.

— На складе слева от того, в котором держат тигра, пьют лимонад, значит, на втором складе, где держат медведя, пьют лимонад.

— Тот, кто пьет сок, разводит пуму, значит, на третьем складе, где едят пельмени, пьют сок.

Итак, ответы на вопросы:

— Кто ест пиццу? Никто не ест пиццу, так как в условии нет информации о том, кто ест пиццу.

— Кто держит пуму? Тот, кто пьет сок, разводит пуму, значит, пуму держат на третьем складе, где пьют сок.

**Уточнение.** А если представить что пиццу ест тот, кто не ест ни хлеб, ни пельмени. Также стоит назвать, кто именно по профессии ест пиццу и кто держит пуму.

Из условий задачи и предположения о том, что пиццу ест тот, кто не ест ни хлеб, ни пельмени, можно сделать следующие выводы:

— Второй склад зеленый, потому что там работает водитель.

— Первый склад красный, так как водитель работает слева от красного склада.

— Третий склад белый, так как центральный склад, где едят хлеб, занят инженером.

— На втором складе, где пьют воду и держат медведя, едят пиццу.

— На третьем складе, где едят пельмени и пьют сок, держат пуму.

Итак, по профессиям:

— Пиццу ест водитель.

— Пуму держит инженер.

Из данного ответа модели следует, что логика прослеживается крайне плохо, практически в каждом утверждении содержится ошибка, первоначально особенность задачи не была понята, в следствии чего ответы были даны очень абстрактно. После подробного разъяснения условий ответы были даны в верном формате и даже угадан ответ с пиццей.

Данной модели также была представлена для решения исходная задача Эйнштейна, но она справилась с ней довольно средне: неправильно ответила на один из вопросов и не поняла особенность задачи. Это примечательно, учитывая, что gpt-2 успешно справилась с исходной задачей.

Модель	Правильность ответа про пиццу	Правильность ответа про пуму	Изменения после второго шанса	Понимание особенности задачи
claude-3-opus-20240229	Да	Нет	Нет	Есть, но не всегда
command-r	Нет	Нет	Нет	В целом да
mistral-large-2402	Да	Да	Нет	Да
llama-2-70b-chat	Нет	Нет	Нет	Нет
llama-3-70b-instruct	Нет	Да	Да	В целом да
gpt2-chatbot	Да	Да	Да	В целом да
gpt-3.5-turbo-0125	Да	Нет	Да	В целом да

Проанализировав все вышеперечисленные ответы, можно сделать несколько выводов.

1. Большинство языковых моделей способно логически мыслить и строить логические цепочки, но зачастую они путаются при большом количестве фактов и задачу Эйнштейна решить не могут, однако при некотором количестве уточнений и у некоторых наблюдается прогресс. Это говорит о том, что языковые модели способны к обучению на подобных задачах [5, 6].

2. Модель GPT-2 верно решила задачу и справилась с оригинальной большой задачей в отличие от GPT-3.5, что очень странно и может говорить о том, что GPT-2 обучали на данной задаче.

3. Модель LLaMA3 справилась с задачей заметно лучше, чем LLaMA2, хотя в целом модели данного семейства не очень сильно обладают логикой.

На основе проведенного анализа и сравнения возможностей различных языковых моделей в решении комплексных логических задач, таких как задача Эйнштейна, можно сделать следующие заключения.

1. Языковые модели, несмотря на свою способность к формированию логических цепочек, испытывают трудности при работе с большим объемом

фактов. Это подчеркивает важность разработки механизмов для улучшения их понимания структуры и связей между данными [7, 8].

2. Наблюдаемый прогресс в решении задач с добавлением уточнений и подсказок указывает на возможности обучения моделей на специфических примерах и их адаптации к изменяющимся условиям задач.

3. Модель GPT-2, несмотря на свою устаревшую архитектуру по сравнению с GPT-3.5, показала лучшие результаты в решении конкретной задачи. Это может свидетельствовать о целенаправленной подготовке или оптимизации модели под определенные типы задач. В то же время более высокие результаты LLaMA3 по сравнению с LLaMA2 подчеркивают прогресс в развитии моделей семейства LLaMA, особенно в аспектах логического мышления и обработки комплексных запросов.

4. Существующие различия в эффективности решения задач между моделями разных поколений и архитектур подчеркивают необходимость продолжения исследований в этой области. Важно не только улучшать существующие модели, но и разрабатывать новые подходы и технологии для обработки и анализа данных [9, 10].

Таким образом, дальнейшее изучение и оптимизация языковых моделей для повышения их логических способностей и обучаемости служит ключевым моментом для расширения их применения в различных областях.

## Литература

- [1] Лосев А.Ф. *Введение в общую теорию языковых моделей. История лингво-философской мысли*. Москва, УРСС, 2004, 296 с.
- [2] Ситжалилов А.Л., Абдураманов З.Ш. Языковые модели для приложений генеративного искусственного интеллекта. *Информационно-компьютерные технологии в экономике, образовании и социальной сфере*, 2024, № 1 (43), с. 68–74. EDN: IAHNLA.
- [3] Лебедев А.С., Марищук Б.В. Технологии генерации текста на основе машинного обучения. *Инженерные технологии*, 2023, № 2 (2), с. 30–34. EDN: SJAPYH.
- [4] Луценко Е.В. *Инженерия знаний и интеллектуальные системы*. Краснодар, Эйдос, 2020, 642 с. <https://doi.org/10.13140/RG.2.2.28085.91364>
- [5] Rogozin O.V. Решение задачи Эйнштейна на основе фреймовой модели представления знаний. *Школьные технологии*, 2010, № 3, с. 151–156.
- [6] Нечаев Н.П., Rogozin O.V. Реализация механизма поддержки принятия решений с использованием фреймовой модели на примере задачи Эйнштейна. *Новые информационные технологии в автоматизированных системах*, 2007, № 10, с. 87–97.

- [7] Мамедова Л.Э., Иванова Л.Н., Алтаев Е.С. Основные аспекты технологии искусственного интеллекта. *Известия вузов ЭфиУП*, 2023, № 3 (57). <https://doi.org/10.6060/ivecofin.2023573.656>
- [8] Горев С.В. Исследование методов и алгоритмов искусственного интеллекта при определении стоимости произведений искусства. *Известия вузов ЭфиУП*, 2022, № 4 (54). <https://doi.org/10.6060/ivecofin.2022544.622>
- [9] Бяшимов П., Какабаева М., Довлетгельдыев Ы., Ораев М. Строительство и искусственный интеллект: преимущества и недостатки. *Символ науки*, 2024, № 4-1-2, с. 32–34.
- [10] Мамиева О.Г., Байджанова Г.Н. Инновационные подходы к обеспечению информационной безопасности в эпоху цифровизации. *Символ науки*, 2024, № 4-1-2, с. 55–57.

*Поступила в редакцию 03.03.2025*

**Томская Ульяна Павловна** — студентка кафедры «Компьютерные системы и сети», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

**Захаров Михаил Александрович** — старший преподаватель кафедры «Компьютерные системы и сети», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

**Ссылку на эту статью просим оформлять следующим образом:**

Томская У.П., Захаров М.А. Анализ логических способностей языковых моделей на примере решения задачи Эйнштейна. *Политехнический молодежный журнал*, 2025, № 04 (99). URL: <https://ptsj.bmstu.ru/catalog/icec/sacip/1058.html>

## ANALYSIS OF THE LOGICAL ABILITIES OF LANGUAGE MODELS USING THE EXAMPLE OF SOLVING THE EINSTEIN PROBLEM

**U.P. Tomskaya**

tomskayaup@student.bmstu.ru

**M.A. Zakharov**

maz@bmstu.ru

*Bauman Moscow State Technical University, Moscow, Russian Federation*

The article examines various approaches to solving the Einstein problem used by language models and compares their effectiveness. The authors analyze the results of the models, identifying their strengths and weaknesses in the context of logical thinking. The article provides examples of solving the Einstein problem using language models from different families — claude, command-r, mistral, llama-2, llama-3, gpt2, gpt-3.5. For each model, an analysis of logical abilities is performed in solving this problem. In the end, the results of solving the problem by different models are summarized and conclusions and observations are made about how each of the models is able to reason.

**Keywords:** language models, artificial intelligence, logical abilities, logical thinking, Einstein's problem, comparative analysis, efficiency

*Received 03.03.2025*

**Tomskaya U.P.** — Student of Computer Systems and Networks Department, Bauman Moscow State Technical University, Moscow, Russian Federation.

**Zakharov M.A.** — Senior lecturer of Department of Computer Systems and Networks, Bauman Moscow State Technical University, Moscow, Russian Federation.

### **Please cite this article in English as:**

Tomskaya U.P., Zakharov M.A. Analysis of the logical abilities of language models using the example of solving the Einstein problem. *Politekhnikheskiy molodezhnyy zhurnal*, 2025, no. 04 (99). (In Russ.). URL: <https://ptsj.bmstu.ru/catalog/icec/sacip/1058.html>