

Введение в LLM и Prompt Engineering с RAG

Сравнение ответов на вопросы из набора данных MMLU

<div><div>deepseek-r1:1.5b</div><div><pre>PS C:\Users\Danya> ollama run deepseek-r1:1.5b <>> Oxygenated blood is carried to the heart by which of the following structures? ... Aorta ... Carotid arteries ... Inferior vena cava ... Pulmonary veins <think> Okay, so I need to figure out what structure carries oxygenated blood to the heart. Let me start by recalling some basic physi logy. The heart pumps blood throughout the body, right? And the blood that goes into the heart is called pulmonary blood because it's part of the lungs and the blood that delivers oxygen to the tissues in the lungs. Now, oxygenated blood needs to go through several vessels before reaching the heart. I think it first travels from the airways into the left ventricle, which is where the pulmonary arteries branch off. Then, those pump blood into the lungs, where they meet the pulmonary veins later on. The pulmonary veins carry the oxygenated blood back do wn the legs. But wait, the question is about oxygenated blood going to the heart. So once the blood comes back from the legs, it needs to go through more vessels. It goes up the aorta into the right and left ventricles of the heart. Then, from there, the blood travels down into the carotid arteries and inferior vena cava. The superior pulmonary carries oxygenated blood from the lungs back to the heart. So putting it all together, oxygenated blood g oes from the heart (aorta) to the carotid and inferior vena cava, then back up through the pulmonary veins, into the left atrium, then further down into the left ventricle where it meets the aorta. Looking at the options given: Aorta is correct because it's the main artery supplying the heart. The other options are not the final part of the journey from the heart to the lungs. So the answer should be the aorta. </think> The oxygenated blood travels from the heart via the <<aorta>> into the carotid arteries and inferior vena cava, then back throu gh the pulmonary veins to the left atrium, and ultimately reaches the left ventricle where it loops around to join the aorta again. Answer: Aorta</pre></div></div> <div><p>Окончательный ответ дан неправильно, но в процессе рассуждений упомянут правильный ответ: "... oxygenated blood goes from heart to ... , then back up through the pulmonary veins ...". Рассуждения используют достоверные сведения о человеческом организме. На основе рассуждений модель делает вывод, что верный ответ - аорта, поскольку она "главная артерия, снабжающая сердце". Обосновывает этот ответ модель тем, что остальные варианты не являются конечными составляющими кровообращения от сердца до лёгких. Модель также не считает левое предсердие частью сердца.</p></div>	<div><div>llama3.2:latest</div><div><pre>PS C:\Users\Danya> ollama run llama3.2 pulling manifest ... 2.6 GB pulling 606ba3fc5ff... 100% 1.4 MB pulling 966de95ca8a6... 100% 7.2 MB pulling f6cb46a9e9da... 100% 6.6 MB pulling a76ff7e576d9... 100% 96 B pulling 56b48bd077a5... 100% 561 B pulling 56b48bd077a5... 100% verifying sha256 digest writing manifest success <>> Oxygenated blood is carried to the heart by which of the following structures? ... Aorta ... Carotid arteries ... Inferior vena cava ... Pulmonary veins The correct answer is A) Aorta. Oxygenated blood from the lungs is carried back to the heart through the pulmonary veins. The pulmonary veins then empty into the left atrium of the heart. From there, the oxygenated blood flows through the left ventricle and is pumped out of the heart through the aortic valve into the aorta, which carries it to the rest of the body. The other options are not correct because: • Carotid arteries (B) carry oxygenated blood from the aorta to the brain. • Inferior vena cava (C) carries deoxygenated blood from the lower body back to the right atrium of the heart. • Pulmonary veins (D) carry oxygenated blood from the lungs to the left atrium, not directly to the heart.</pre></div></div> <div><p>Окончательный ответ дан неправильно, но в процессе рассуждений упомянут правильный ответ: "Oxygenated blood from lungs is carried back to the heart through the pulmonary veins". Дальнейшие рассуждения используют достоверные сведения о человеческом организме. Модель отбрасывает два заведомо неверных варианта и верный вариант, так как модель не считает левое предсердие частью сердца. Таким образом, модель получает единственный вариант - аорта.</p></div>
<div><div>deepseek-r1:1.5b - ></div><div><p>A program is expressed in a programming language. Which of the following is true of the program?The program can also be expressed as binary code, but will be more easily understood by humans when ...The program can also be expressed as binary code, which will reduce the likelihood of errors.The program cannot be expressed as binary code, because binary code can only be used to represent da...Some parts of the program can be expressed as binary code, but operations must be expressed using a...</p><div><div>deepseek-r1:1.5b</div><div><p>Thought for 2 seconds</p><p>The program does not require all parts to be expressed in binary code for better understanding by humans. Instead, some components can remain in human-readable form, while others are converted to binary code. This approach improves readability and reduces the likelihood of errors.</p><p>Answer: The correct statement is that some parts of the program can be expressed as binary code while other parts remain in human-readable form, enhancing clarity and reducing errors.</p></div></div></div><div><p>Модель даёт верный, но неполный ответ на поставленный вопрос. Модель понимает структуру вопроса, но не предоставляет все необходимые рассуждения, на основании которых выбрано утверждение. Модель не дополняет выбранное утверждение.</p></div></div>	<div><div>llama3.2:latest - ></div><div><p>A program is expressed in a programming language. Which of the following is true of the program?The program can also be expressed as binary code, but will be more easily understood by humans when ...The program can also be expressed as binary code, which will reduce the likelihood of errors.The program cannot be expressed as binary code, because binary code can only be used to represent da...Some parts of the program can be expressed as binary code, but operations must be expressed using a...</p><div><div>llama3.2:latest</div><div><p>It'll help you with that.</p><p>The correct answers are:</p><ol style="list-style-type: none">The program can also be expressed as binary code, but will be more easily understood by humans when it is translated into an intermediate language (e.g., assembly language or bytecode).The program can also be expressed as binary code, which will reduce the likelihood of errors.None of the above options are correct. Binary code can represent any program, regardless of the programming language used to write it.Some parts of the program can be expressed as binary code, but operations must be expressed using a higher-level language (e.g., C or Python).<p>In general, programs can be expressed in various forms:</p><ul style="list-style-type: none">Source code: written by humans using a high-level languageBinary code: machine-specific code that is directly executed by the computer's processorAssembly code: translated from source code into binary code using an assemblerIntermediate language (e.g., bytecode): used as an intermediate form between source code and binary code.<p>Each of these forms has its own advantages and disadvantages, depending on the specific use case.</p></div></div></div><div><p>Модель даёт неправильный ответ на поставленный вопрос. Модель не понимает структуру вопроса. В ответе модель не даёт свои рассуждения, а предоставляет объяснение различных видов представления программы.</p></div></div>

You

What is the sum of all positive integer values of n such that n^2 is a factor of 1200?

42
12
36
39

Assistant bartowski/deepseek-r1-distill-qwen-1.5b

Thoughts

> Thought for 24.33 seconds

To solve the problem of finding the sum of all positive integer values of n such that n^2 is a factor of 1200, we start by factorizing 1200 into its prime factors:

$$1200 = 2^4 \times 3^1 \times 5^2$$

We need to find all positive integers n such that n^2 divides 1200. This means the exponents in the prime factorization of n^2 must be less than or equal to those in 1200.

For each prime factor:

- For prime 2: The exponent in n^2 can be at most 4, so the exponent in n can be 0, 1, or 2.
- For prime 3: The exponent in n^2 must be even and less than or equal to 1. Since it must be an even number, the only possible value is 0.
- For prime 5: The exponent in n^2 can be at most 2, so the exponent in n can be 0 or 1.

Thus, n can be written as $2^a \times 3^c \times 5^d$ where:

- $a \in \{0, 1, 2\}$
- $c = 0$ (since the exponent of 3 in n^2 must be even and at most 1)

This gives us possible values for n :

- When $a = 0, d = 0$: $n = 1$
- When $a = 0, d = 1$: $n = 5$
- When $a = 1, d = 0$: $n = 2$
- When $a = 1, d = 1$: $n = 10$
- When $a = 2, d = 0$: $n = 4$
- When $a = 2, d = 1$: $n = 20$

Summing these values:
 $1 + 5 + 2 + 10 + 4 + 20 = 42$

Thus, the sum of all positive integer values of n is:

A 42

Модель даёт правильный ответ на поставленный вопрос. Модель понимает структуру вопроса. В ответе модель даёт свои аналитические рассуждения, где представляет задачу в более комплексном виде, предоставляя обоснование проделанных шагов.

You

What is the sum of all positive integer values of n such that n^2 is a factor of 1200?

42
12
36
39

Assistant llama-3.2-3b-instruct-gpt_8

The prime factorization of 1200 is:

$$1200 = 2^4 \cdot 3 \cdot 5^2$$

To find the values of n , we need to find all combinations of the prime factors that can be squared.

The possible values are 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, and 24.

Only these 16 numbers will work because if a number is not one of these, then its square will have at least one prime factor with an exponent greater than one and therefore it can't divide into 1200.

Since the sum of all positive integer values of n such that n^2 divides into 1200 is

$$1 + 2 + 3 + 4 + 5 + 6 + 8 + 9 + 10 + 12 + 15 + 16 + 18 + 20 + 24 = 144,$$

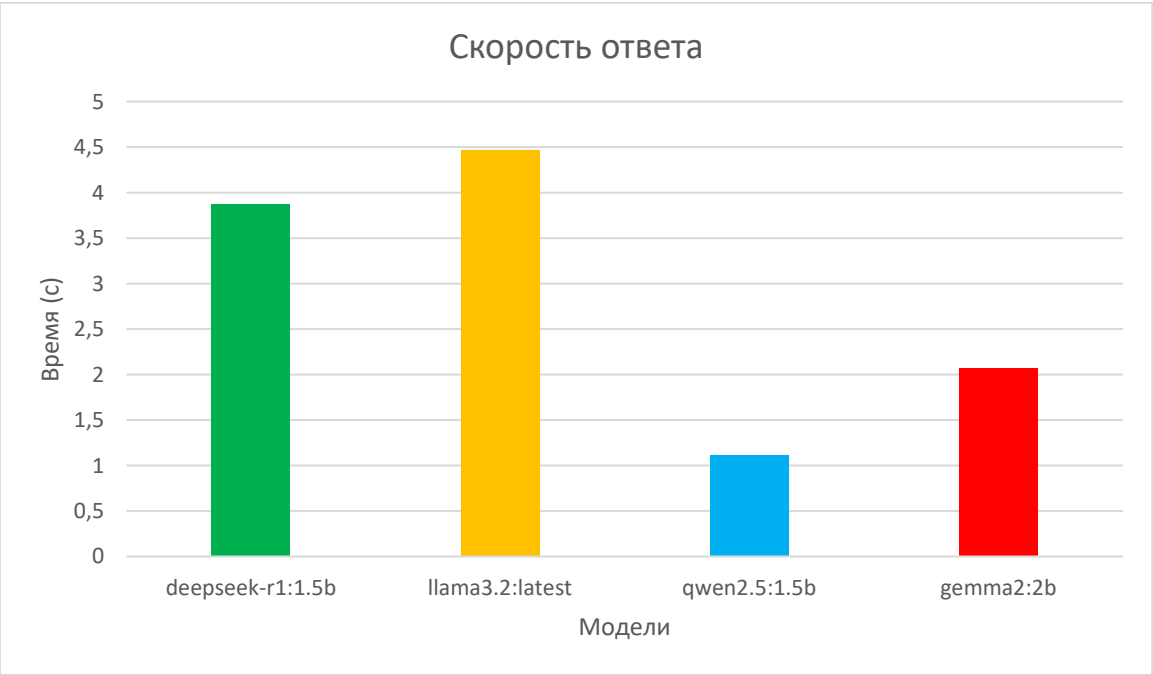
then our answer is

144

112.07 tok/sec • 203 tokens • 0.14s to first token • Stop reason: EOS Token Found

Модель даёт неправильный ответ на поставленный вопрос. Модель понимает структуру вопроса. В ответе модель даёт свои рассуждения, но допускает логические ошибки.

Сравнение ответов на вопрос по информации, полученной веб-скрапингом



(Ответы находятся в репозитории по пути homework01/history/chat_history.txt)

Все четыре LLM (deepseek-r1:1.5b, llama3.2:latest, qwen2.5:1.5b, gemma2:2b) эффективно отвечают на вопрос пользователя о вреде сахара, опираясь на предоставленный контекст. Однако они различаются по своему подходу, уровню детализации и форматированию.

deepseek-r1:1.5b: кратко и прямо отвечает на вопрос, обобщая рекомендации ВОЗ и общий вред высокого потребления сахара, включая кардиометаболические проблемы. Он явно упоминает предоставленный контекст. Форматирование простое.

Llama3.2:latest: предоставляет более структурированный и подробный ответ, используя заголовки и маркеры. Он охватывает общие эффекты, противоречия, связанные с финансированием сахарной промышленности, потенциальные связи с раком (отмечая при этом неубедительность доказательств) и краткое заключение. Он включает более широкий спектр информации из предполагаемого более широкого контекста (хотя и не все явно указано в истории чата последнего хода). Форматирование хорошо организовано.

qwen2.5:1.5b: ответ очень краткий и фокусируется на значительном вреде сахара, его связи с кардиометаболическими проблемами, ожирением, диабетом и метаболическим синдромом, и прямо цитирует рекомендации ВОЗ. Это очень прямолинейный ответ.

gemma2:2b: ответ также довольно краткий и констатирует общий вред и связь с кардиометаболическими проблемами. Затем он конкретно выделяет связь между подслащенными сахаром напитками и увеличением веса и повторяет их общее негативное воздействие. Он явно упоминает использование предоставленного контекста. Форматирование четкое.

Сильные и слабые стороны каждого ответа

deepseek-r1:1.5b

Сильные стороны: Краткий прямой ответ на вопрос, явное использование предоставленного контекста.

Слабые стороны: Менее детализированный ответ, чем некоторые другие.

llama3.2:latest

Сильные стороны: Очень подробный и хорошо структурированный ответ, охватывающий множество аспектов темы (включая влияние промышленности и риск рака).

Слабые стороны: Часть представленной информации (например, подробный раздел о финансировании сахарной промышленности и риске рака) не была явно представлена в истории чата последнего хода, что предполагает использование более широкой базы знаний или предыдущих ходов, не показанных здесь.

qwen2.5:1.5b

Сильные стороны: Чрезвычайно краткий ответ, прямо затрагивающий ключевые вредные факторы и рекомендации.

Слабые стороны: Менее детализированный ответ, чем другие.

gemma2:2b

Сильные стороны: Краткость, четкое изложение основных моментов, явные ссылки на контекст с примерами.

Слабые стороны: Менее всеобъемлющий ответ, чем у llama3.2:latest.

LLM	Полнота	Точность (на основе контекста)	Четкость и краткость	Использование контекста	Форматирование и структура
deepseek-r1:1.5b	Средняя	Высокая	Высокая	Высокое	Простое
llama3.2:latest	Высокая	Вероятно высокая (предполагает более широкий контекст)	Средняя	Среднее	Хорошее
qwen2.5:1.5b	Низкая	Высокая	Высокая	Высокое	Простое
gemma2:2b	Средняя	Высокая	Высокая	Высокое	Чёткое