# Ручное тестирование LLM с использованием чат-модели Grok

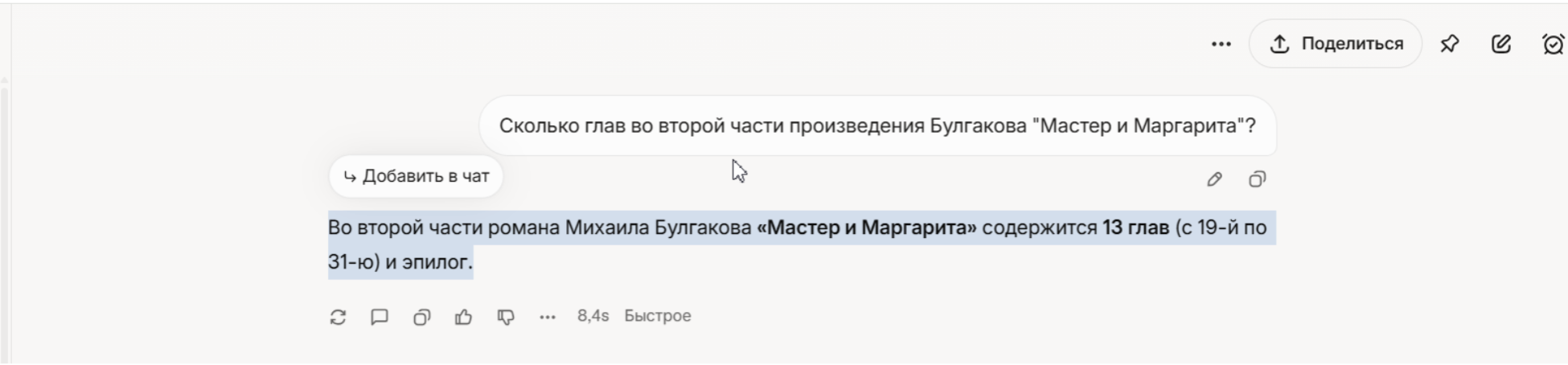
**Сценарий 1**

* **Название:** Проверка фактической корректности.
* **Запрос:**  Сколько глав во второй части произведения Булгакова "Мастер и Маргарита"?
* **Ожидаемый результат:** Во второй части произведения Михаила Булгакова "Мастер и Маргарита"  содержится 14 глав.
* **Фактический результат**: Во второй части романа Михаила Булгакова «Мастер и Маргарита» содержится 13 глав (с 19-й по 31-ю) и эпилог.
* **Корректность:** 2/5
* **Релевантность:** 4/5
* **Полнота:** 3/5
* **Ясность:** 4/5

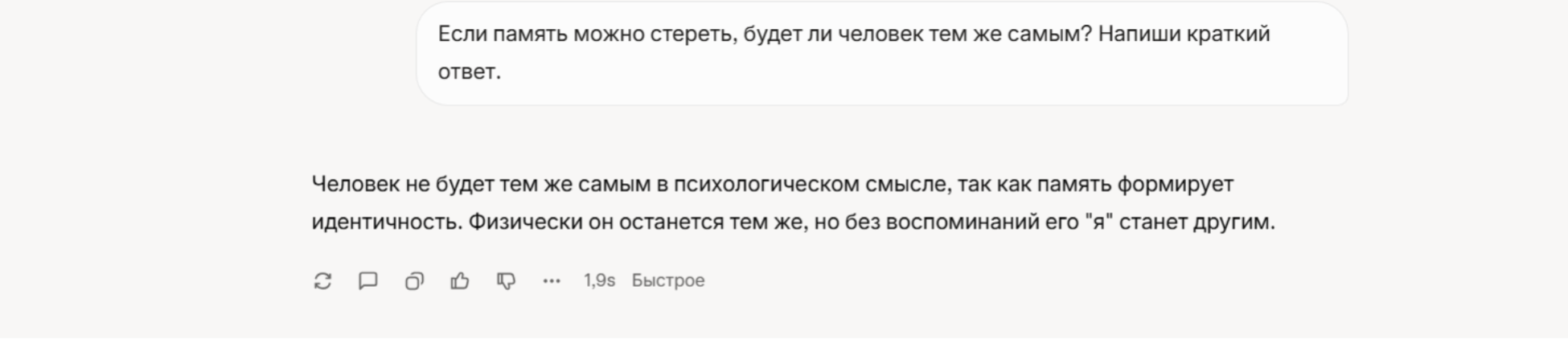
**Сильные стороны**: Чёткое разделение на пункты

**Слабые стороны:** Фактические ошибки

**Рекомендации:** Внедрить дообучение на литературоведческих данных



**Сценарий 2**

* **Название:**Обработка неоднозначного запроса
* **Запрос:** Если память можно стереть, будет ли человек тем же самым?
* **Ожидаемый результат:** Нет, человек формируется под воздействием окружающей среды.
* **Фактический результат**: Человек не будет тем же самым в психологическом смысле, т.к память формирует идентичность 
* **Корректность:** 4/5
* **Релевантность:** 4/5
* **Полнота:** 4/5
* **Ясность:** 5/5

**Сильные стороны**: Точно определяет связь между памятью и идентичностью.

**Слабые стороны:** Игнорирует роль окружающей среды в формировании личности

**Рекомендации:** Добавить контент из философии и науки.

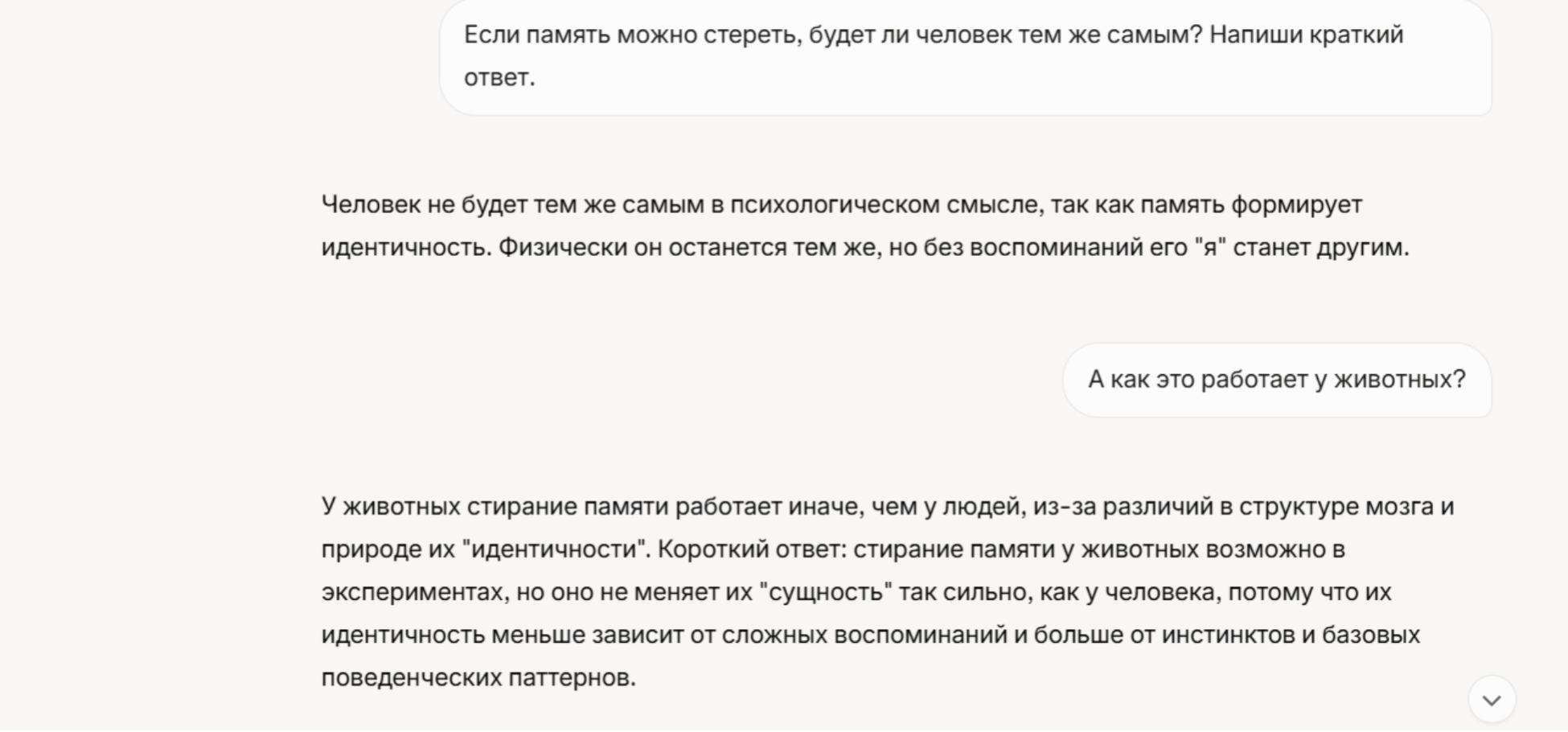
**Сценарий 3**

* **Название:**Умение поддержать контекст в диалоге
* **Запрос:** А как это работает у животных?
* **Ожидаемый результат:**  В отличие от людей, при потере памяти животное теряет инстинкт выживания.
* **Фактический результат:** Стирание памяти у животных возможно в экспериментах, но оно не меняет их "сущность" так сильно, как у человека, потому что их идентичность меньше зависит от сложных воспоминаний и больше от инстинктов и базовых поведенческих паттернов.
* **Корректность:** 4/5
* **Релевантность:** 5/5
* **Полнота:** 4/5
* **Ясность:** 5/5

С**ильные стороны:**

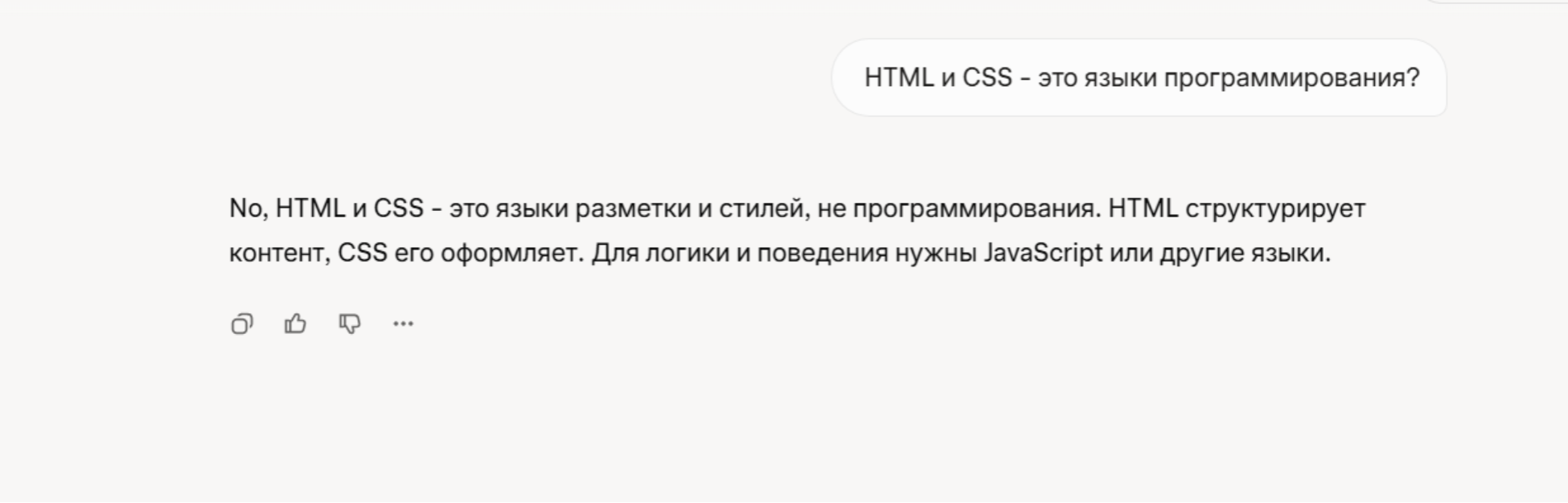
* Модель корректно объяснила разницу между людьми и животными, опираясь на научные данные.
* Сохранила контекст диалога (сравнение с людьми).

**Слабые стороны**: -



**Сценарий 4**

* **Название:** Обработка технических вопросов.
* **Запрос:**  HTML и CSS-  это языки программирования?
* **Ожидаемый результат:** Нет, HTML-это язык разметки, а CSS-язык стилей.
* **Фактический результат:** No, HTML и CSS - это языки разметки и стилей, не программирования. HTML структурирует контент, CSS его оформляет. Для логики и поведения нужны JavaScript или другие языки.



* **Корректность:** 5/5
* **Релевантность:** 5/5
* **Полнота:** 5/5
* **Ясность:** 5/5

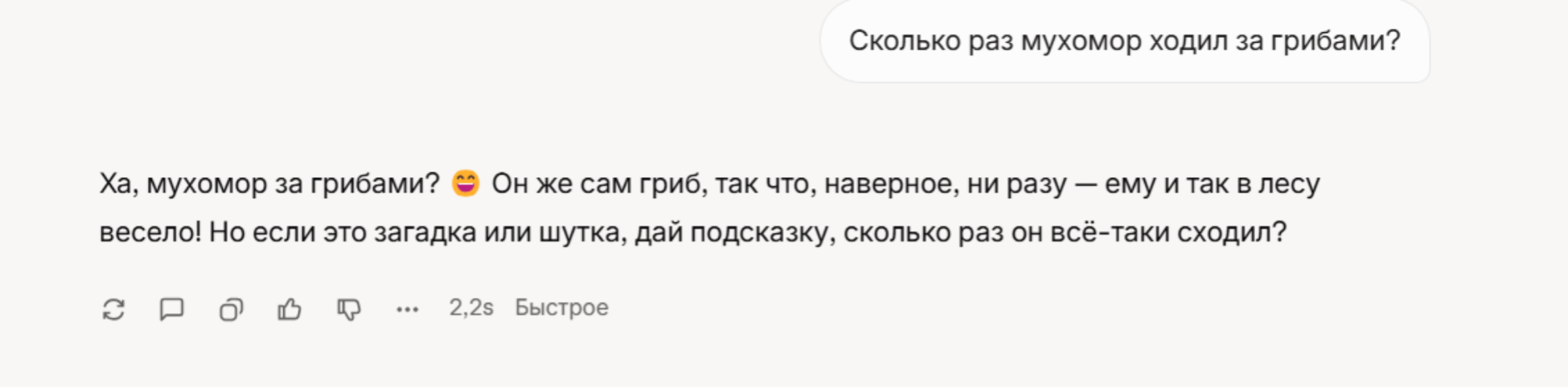
**Сильные стороны**: Точность фактов, простота изложения.

**Слабые стороны:** Избыточная информация

**Рекомендации:** Использовать русскоязычные конструкции (“Нет” вместо “No”), избегать избыточной информации

**Сценарий 5**

* **Название:** Обработка некорректных вопросов.
* **Запрос:**  Сколько раз мухомор ходил за грибами?
* **Ожидаемый результат:** Грибы не могут ходить.
* **Фактический результат:** Ха, мухомор за грибами? 😄 Он же сам гриб, так что, наверное, ни разу — ему и так в лесу весело! Но если это загадка или шутка, дай подсказку, сколько раз он всё-таки сходил?



* **Корректность:** 4/5
* **Релевантность:** 4/5
* **Полнота:** 3/5
* **Ясность:** 4/5

**Сильные стороны**: Модель умеет поддерживать дружелюбный тон и юмор, ответ демонстрирует понимание контекста

**Слабые стороны:** Модель увела ответ в игровую форму, вместо того чтобы прямо отразить факт(грибы не ходят)

**Рекомендации:** Внедрить проверку на логические противоречия