

# Шкала оценки: почему умеренная шкала (1-10) работает лучше, чем узкая (1-3)

Исследования в области промпт-инжиниринга выявили интересную закономерность: для более точных и полезных результатов лучше использовать умеренные шкалы оценки (например, от 1 до 10), чем слишком узкие (от 1 до 3). Давайте разберемся почему это так и как мы можем использовать это знание.

## Основные принципы

1. **Более тонкая градация:** Шкала 1-10 предоставляет больше промежуточных значений, позволяя выразить более нюансированную оценку.
2. **Больше пространства для различий:** Расширенная шкала позволяет лучше различать близкие, но не идентичные случаи.
3. **Лучшая калибровка:** Модели AI способны лучше калибровать свои оценки, когда у них есть более широкий диапазон значений.
4. **Привычность для людей:** Десятибалльная шкала интуитивно понятна большинству людей, что делает взаимодействие более естественным.
5. **Возможность группировки:** Шкала 1-10 легко группируется в логические категории (например, 1-3: слабо, 4-6: средне, 7-10: сильно), сохраняя при этом детализацию.

## Исследования, подтверждающие эффективность

Согласно данным из базы исследований:

1. **HPSS (Эвристическая стратегия поиска подсказок):** Исследование показало, что умеренные шкалы оценки (1-10) с четкими диапазонами для разных уровней качества дают более последовательные и обоснованные результаты.
2. **RevisEval:** Данное исследование использует шкалу от 1 до 10 для оценки параметров, таких как точность, полнота и ясность, что повышает точность оценки на 2-6%.
3. **Улучшение вывода LLM как судьи:** Исследователи обнаружили, что можно получать более детальные результаты, используя шкалы с большим диапазоном (например, от 1 до 100 вместо 1-5), позволяя модели выражать небольшие различия в качестве.

4. **Исследование личностей AI:** При оценке человекоподобности рассуждений вместо стандартной 3-балльной использовалась 6-балльная шкала, что позволило получить более детальное и близкое к человеческому распределение оценок.

## Практические примеры

### Пример 1: Оценка аргументации в эссе

Оцени качество аргументации в предоставленном эссе по шкале от 1 до 10, где:

- 1-3: слабая аргументация
- 4-6: средняя аргументация
- 7-10: сильная аргументация

Критерии оценки:

- Логическая связность аргументов
- Использование доказательств и примеров
- Рассмотрение контраргументов
- Убедительность общей позиции

### Пример 2: Оценка ответа на вопрос

Оцени ответ по шкале от 1 до 10 по следующим критериям:

- Точность (насколько информация корректна)
- Полнота (насколько охвачены все аспекты вопроса)
- Ясность (насколько понятно объяснение)

### Пример 3: Сравнение вариантов решения

Оцени каждый из предложенных вариантов решения по шкале от 1 до 10.

Вместо выбора одного конкретного балла, укажи диапазон вероятных оценок (например, "7.2-7.8").

Затем выполни сравнение вариантов, указав, насколько один вариант превосходит другой.

## Почему это работает

1. **Психологический фактор:** Когда мы используем слишком узкую шкалу (1-3), модель часто склоняется к средним значениям, избегая крайностей. Шкала 1-10 позволяет более свободно использовать весь диапазон оценок.

2. **Больше информации для обучения:** Расширенные шкалы содержат больше информации, что позволяет моделям лучше дифференцировать оценки и обосновывать своё решение.
3. **Снижение позиционных смещений:** Исследования показывают, что расширенные шкалы помогают снизить смещение в оценках и делают их более объективными.
4. **Возможность выражения неопределенности:** Широкие шкалы позволяют выражать неуверенность в оценке через указание диапазона (например, 7-8 баллов), что более реалистично отражает человеческий процесс оценки.
5. **Совместимость с многокритериальной оценкой:** Более детальные шкалы лучше работают при оценке нескольких аспектов одновременно (точность, полнота, ясность и т.д.).

## Практические рекомендации по применению

1. **Определяйте значения для диапазонов:** Обязательно указывайте, что означают разные части шкалы (например, 1-3: слабо, 4-6: средне, 7-10: сильно).
2. **Разбивайте оценку на критерии:** Используйте шкалу 1-10 для оценки различных аспектов, а не только для общего впечатления.
3. **Запрашивайте обоснование:** Попросите AI объяснить, почему была присвоена конкретная оценка, это улучшает качество результатов.
4. **Для сравнения нескольких вариантов:** Попросите указать не только абсолютные оценки, но и относительную разницу между вариантами.
5. **Для тонких различий:** Если вам нужно выявить очень небольшие различия, можно использовать даже более широкие шкалы (1-100) или дробные значения.

Умеренные шкалы оценки — это простой, но эффективный инструмент, который позволяет получать более точные, нюансированные и полезные результаты при взаимодействии с AI, будь то оценка текста, сравнение вариантов или анализ контента.