

Chain-of-Verification (CoV) — Цепочка Верификации

Chain-of-Verification (CoV) — это эффективный метод улучшения точности языковых моделей через механизм самопроверки, в котором модель проверяет и исправляет свои собственные ответы.

Принципы для применения в промптах

1. Декомпозиция процесса решения

- Разбивайте сложную задачу на последовательные логические шаги
- Для каждого шага запрашивайте четкий промежуточный результат

2. Самопроверка на основе вопросов

- После получения первоначального решения, попросите модель генерировать проверочные вопросы
- Вопросы должны быть критическими и направленными на выявление потенциальных ошибок

3. Итеративное улучшение решения

- На основе ответов на проверочные вопросы, модель корректирует своё решение
- Запрашивайте пояснения к внесённым изменениям

4. Прозрачная фиксация корректировок

- Просите модель явно указывать, какие коррекции были внесены и почему
- Сохраняйте как исходное, так и исправленное решение для сравнения

5. Структурированный формат проверки

- Используйте чёткую структуру для процесса верификации
- Разделяйте исходные данные, выводные результаты и области неопределённости

Практический пример промпта

Решение задачи с использованием Chain-of-Verification

Задача:

[Описание задачи]

Инструкции:

1. Сначала реши задачу, последовательно выполняя следующие шаги:

- Шаг 1: [описание шага]
- Шаг 2: [описание шага]
- ...
- Шаг N: [описание шага]

2. После получения первоначального решения, сформулируй 3-5 критических вопросов для проверки:

- Какие потенциальные ошибки могут быть в решении?
- Что можно упустить или неправильно понять?
- Какие альтернативные подходы следует рассмотреть?

3. Ответь на каждый проверочный вопрос:

- Дай подробный и честный ответ на каждый вопрос
- Укажи, выявил ли вопрос проблему в первоначальном решении

4. Внеси необходимые коррективы:

- На основе ответов на вопросы исправь своё первоначальное решение
- Четко выдели, какие изменения были внесены и почему

5. Представь финальное решение:

- Дай окончательное исправленное решение
- Укажи, какие коррективы были внесены и почему они важны

Как это работает и почему это эффективно

- Преодоление когнитивных предубеждений** Chain-of-Verification работает, заставляя модель смотреть на собственный ответ критически и с другой перспективы. Это помогает преодолеть "подтверждающее предубеждение" - тенденцию поддерживать уже сделанные выводы.
- Снижение ошибок через самопроверку** Метод значительно снижает вероятность ошибок в вычислениях и логических рассуждениях, так как

модель выступает сама себе оппонентом, активно ища слабые места в своём первоначальном ответе.

3. **Структурирование процесса мышления** CoV предоставляет чёткую структуру для обработки информации, что особенно полезно для сложных задач с множеством переменных или этапов, где вероятность ошибки выше.
4. **Повышение прозрачности рассуждений** Весь процесс верификации доступен пользователю, что повышает доверие к результату и позволяет оценить качество рассуждений модели.
5. **Итеративное улучшение качества** Каждая итерация проверки и коррекции повышает качество ответа, причём этот процесс можно повторять несколько раз для особенно сложных задач.

Этот метод особенно эффективен для задач с высокими требованиями к точности: финансовых расчётов, юридических анализов, медицинских диагностик, и любых других сценариев, где ошибки могут иметь серьёзные последствия. Исследования показывают, что применение CoV может значительно снизить частоту ошибок и "галлюцинаций" в ответах языковых моделей.