

HoT: Выделенная цепочка размышлений для ссылки на поддерживающие факты из ВХОДНЫХ ДАННЫХ

Дата: 2025-03-04 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2503.02003>

Рейтинг: 85

Адаптивность: 90

Ключевые выводы:

Исследование представляет новый метод промптинга LLM под названием Highlighted Chain of Thought (HoT), который позволяет моделям генерировать ответы с XML-тегами, связывающими факты в ответе с фактами из исходного вопроса. Основная цель - уменьшить галлюцинации и улучшить верифицируемость ответов LLM. Результаты показывают, что HoT повышает точность LLM на различных задачах и помогает пользователям быстрее проверять ответы.

Объяснение метода:

HoT - это техника промптинга, позволяющая LLM выделять ключевые факты в вопросе и ссылаться на них в ответе. Метод повышает точность ответов на 1.6-2.5%, ускоряет проверку ответов пользователями на 25% и не требует API или дообучения. Легко применим в обычных чатах, работает с различными задачами и моделями, предлагает конкретную методологию создания эффективных примеров.

Ключевые аспекты исследования: 1. **Highlighted Chain of Thought (HoT)** - метод промптинга LLM, который заставляет модели выделять ключевые факты в вопросе XML-тегами, а затем ссылаться на эти факты в ответе с помощью соответствующих тегов, создавая визуальные выделения.

Улучшение точности ответов - исследование показывает, что HoT последовательно повышает точность ответов LLM в среднем на 1.6-2.5 процентных пункта по сравнению с обычным Chain of Thought (CoT) на 17 различных задачах.

Улучшение верификации для пользователей - выделение важных фактов в цепочках рассуждений помогает пользователям на 25% быстрее проверять ответы LLM (47.26 секунд против 62.38 секунд).

Компонентный анализ - исследование показывает, что как повторение вопроса, так и добавление тегов в вопрос и ответ вносят вклад в повышение эффективности

метода HoT.

Метод разработки демонстраций - предлагается подход для создания примеров HoT с помощью LLM, что делает метод более доступным для практического применения.

Дополнение:

Действительно, для работы методов этого исследования **не требуется дообучение или API**. HoT (Highlighted Chain of Thought) - это чистая техника промптинга, которую можно применить в любом стандартном чате с LLM.

Ключевые концепции и подходы для стандартного чата:

Выделение ключевых фактов - можно инструктировать модель выделять важные факты из вопроса в своем ответе, используя маркировку (например, **жирный шрифт**, *курсив* или другие визуальные выделения вместо XML-тегов).

Структура запроса с повторением - исследование показало, что само повторение вопроса с выделением ключевых фактов улучшает точность ответов. Пользователи могут просить модель сначала повторить вопрос с выделением ключевых фактов.

Связывание ответов с фактами - можно просить модель явно указывать, на какие факты из вопроса она опирается при формировании каждой части ответа.

Создание собственных примеров - пользователи могут создавать свои примеры выделения ключевых фактов и их использования в рассуждениях для конкретных типов задач.

Ожидаемые результаты:

Повышение точности ответов - особенно заметно на сложных задачах, требующих точного отслеживания фактов.

Улучшение верифицируемости - ответы становятся более прозрачными, пользователю легче понять, откуда модель берет информацию.

Снижение галлюцинаций - явное связывание с фактами из вопроса снижает вероятность придумывания несуществующей информации.

Ускорение проверки - выделение ключевых моментов делает процесс верификации более быстрым и менее утомительным.

Эта техника особенно ценна для задач, требующих точности и отслеживания множества фактов - от решения математических задач до анализа сложных текстов и логических рассуждений.

Анализ практической применимости: 1. **Highlighted Chain of Thought (HoT)** -

Прямая применимость: Высокая. Пользователи могут сразу применять эту технику в обычных чатах с LLM, добавляя соответствующие инструкции в свои запросы. - Концептуальная ценность: Значительная. Метод помогает лучше понять, как LLM обрабатывают факты из запросов и используют их в ответах. - Потенциал для адаптации: Отличный. Метод может быть адаптирован для различных задач, от математических вычислений до понимания текста.

Улучшение точности ответов Прямая применимость: Средняя. Пользователи получают более точные ответы, но требуется специальное формирование запроса. Концептуальная ценность: Высокая. Демонстрирует важность явной связи между фактами из вопроса и их использованием в ответе. Потенциал для адаптации: Хороший. Принцип выделения ключевых фактов может быть применен в различных контекстах.

Улучшение верификации для пользователей

Прямая применимость: Очень высокая. Пользователи могут сразу применять метод для более быстрой проверки ответов LLM. Концептуальная ценность: Значительная. Показывает, как визуальное выделение влияет на когнитивную нагрузку пользователей. Потенциал для адаптации: Отличный. Может быть адаптирован для различных интерфейсов и задач.

Компонентный анализ

Прямая применимость: Средняя. Знание о вкладе различных компонентов помогает оптимизировать запросы. Концептуальная ценность: Высокая. Демонстрирует важность структуры запроса для качества ответа. Потенциал для адаптации: Хороший. Принципы могут быть применены к другим техникам промптинга.

Метод разработки демонстраций

Прямая применимость: Высокая. Пользователи могут создавать свои собственные примеры для специфических задач. Концептуальная ценность: Средняя. Показывает подход к созданию эффективных примеров. Потенциал для адаптации: Хороший. Метод может быть адаптирован для создания примеров для других техник.

Prompt:

Использование метода Highlighted Chain of Thought (HoT) в промптах для GPT ##
Суть метода HoT Метод HoT предлагает создавать промпты, которые заставляют языковую модель: 1. Переформулировать исходный вопрос с выделением ключевых фактов XML-тегами 2. Генерировать ответ, где каждое утверждение содержит ссылки на выделенные факты из вопроса 3. Это повышает точность ответов и делает их более верифицируемыми

Пример промпта с использованием HoT

[=====] Я хочу, чтобы ты использовал метод Highlighted Chain of Thought (HoT) для

ответа на мой вопрос. Вот как это работает:

Сначала переформулируй мой вопрос, выделяя ключевые факты с помощью XML-тегов ..., ... и т.д. Затем дай подробный ответ, где каждое утверждение в твоём рассуждении будет содержать ссылки на факты из вопроса в формате это утверждение основано на фактах 1 и 2 Убедись, что твой ответ опирается только на информацию из вопроса, чтобы избежать галлюцинаций. Вот мой вопрос: У Анны было 24 яблока. Она отдала 5 яблок Марку и 7 яблок Лизе. Затем она купила еще 12 яблок. Сколько яблок у Анны теперь? [=====]

Как это работает

Улучшение точности: Исследование показало, что HoT повышает точность на +1.60-2.58 процентных пункта в зависимости от типа задачи.

Повышение прозрачности: Модель вынуждена явно связывать свои выводы с конкретными фактами из вопроса, что делает рассуждение более прозрачным.

Ускорение верификации: Пользователи тратят на 25% меньше времени при проверке ответов с выделениями, так как легче проследить, откуда модель берет информацию.

Уменьшение галлюцинаций: Поскольку модель должна ссылаться на конкретные факты из вопроса, это снижает вероятность выдумывания информации.

Когда использовать

Этот подход особенно полезен для: - Арифметических задач - Вопросно-ответных задач с фактическими данными - Задач логического рассуждения - Ситуаций, где важна верифицируемость ответов

Помните, что выделения могут создавать у пользователей ложное чувство уверенности, поэтому важно сохранять критическое мышление при оценке ответов.