# Генерация с поддержкой извлечения на основе ретроактивности доказательств в больших языковых моделях

Дата: 2025-01-07 00:00:00

Ссылка на исследование: https://arxiv.org/pdf/2501.05475

Рейтинг: 65
Адаптивность: 70

## Ключевые выводы:

Исследование представляет новый фреймворк RetroRAG (Retroactive Retrieval Augmented Generation), который решает проблему галлюцинаций в LLM при ответах на сложные многоэтапные вопросы. В отличие от традиционных подходов RAG с однонаправленным рассуждением, RetroRAG использует ретроактивную парадигму, позволяющую пересматривать и корректировать цепочку рассуждений на основе новых доказательств.

## Объяснение метода:

RetroRAG предлагает ретроактивный подход к рассуждениям в LLM, позволяющий пересматривать выводы. Хотя полная реализация технически сложна, концепции разделения доказательств, итеративного улучшения ответов и самосогласованности могут быть адаптированы. Пользователи могут структурировать запросы, разделяя факты и выводы, и применять многошаговые итерации для уточнения ответов.

## Ключевые аспекты исследования: 1. **RetroRAG** - новый подход к извлечению и использованию информации для LLM, основанный на ретроактивной парадигме рассуждений, в отличие от традиционных однонаправленных методов.

**Структура ELLERY** (Evidence-Collation and Discovery) - система, которая собирает, генерирует и обновляет доказательства для построения эффективных цепочек рассуждений, позволяя модели пересматривать свои выводы.

**Двухкомпонентный процесс**: Answerer (генерирует ответы) и ELLERY (управляет доказательствами) - работают итеративно, переоценивая и улучшая ответы.

**Разделение доказательств на исходные и выводимые** - метод позволяет отделять фактические данные от умозаключений, уменьшая галлюцинации.

**Механизм самосогласованности** - оценивает надежность ответов, проверяя их

согласованность при разных температурах генерации.

## Дополнение: Анализируя исследование RetroRAG, можно сделать вывод, что для полной реализации описанных методов действительно требуется API и специальная инфраструктура. Однако многие концепции и подходы можно адаптировать для работы в стандартном чате:

**Ретроактивное мышление** - пользователь может имитировать этот процесс, явно указывая модели пересмотреть предыдущие выводы в свете новой информации: "Давай пересмотрим предыдущее рассуждение, учитывая новый факт X".

**Разделение доказательств** - можно структурировать запрос, явно выделяя "исходные факты" и "выводы из фактов", что поможет модели лучше отделять фактическую информацию от умозаключений.

**Итеративное улучшение** - пользователь может последовательно улучшать ответ через серию уточняющих запросов, каждый раз сохраняя предыдущий контекст.

**Проверка самосогласованности** - можно задать один и тот же вопрос несколькими способами и сравнить ответы для оценки их надежности.

**Поиск недостающей информации** - можно явно спрашивать модель: "Какая дополнительная информация нужна, чтобы ответить на этот вопрос более точно?".

Применяя эти концепции, пользователи могут добиться: - Более точных ответов на сложные многоэтапные вопросы - Снижения "галлюцинаций" модели - Более структурированных и проверяемых рассуждений - Лучшего понимания, как модель пришла к определенному выводу

Таким образом, хотя полная архитектура RetroRAG требует технической реализации, её концептуальные основы могут значительно улучшить взаимодействие с LLM в стандартном чате.

## Анализ практической применимости: **RetroRAG - ретроактивная парадигма рассуждений:** - Прямая применимость: Ограничена, так как требует доступа к API модели и использования нескольких промежуточных шагов. - Концептуальная ценность: Высокая. Идея возврата к предыдущим шагам рассуждения и их пересмотра может быть адаптирована пользователями при формулировании сложных запросов. - Потенциал для адаптации: Значительный. Принцип проверки и пересмотра промежуточных выводов может применяться в обычных диалогах с LLM.

**ELLERY - структура для обработки доказательств:** - Прямая применимость: Низкая для обычных пользователей, так как требует сложной технической реализации. - Концептуальная ценность: Высокая. Разделение на исходные данные и выводы помогает понять, как LLM формируют ответы. - Потенциал для адаптации: Средний. Пользователи могут вручную организовывать информацию схожим образом при работе с LLM.

**Двухкомпонентный итеративный процесс:** - Прямая применимость: Низкая без специальных инструментов. - Концептуальная ценность: Высокая. Помогает понять, что качество ответов LLM можно улучшать через итерации. - Потенциал для адаптации: Высокий. Пользователи могут применять итеративный подход, постепенно уточняя запросы.

**Разделение доказательств:** - Прямая применимость: Средняя. Пользователи могут явно разделять факты и выводы в своих запросах. - Концептуальная ценность: Высокая. Помогает понимать, как структурировать информацию для LLM. - Потенциал для адаптации: Высокий. Этот принцип легко использовать в обычных запросах.

**Механизм самосогласованности:** - Прямая применимость: Низкая без специальных инструментов. - Концептуальная ценность: Высокая. Понимание, что LLM могут давать разные ответы при разных параметрах. - Потенциал для адаптации: Средний. Пользователи могут задавать один вопрос несколькими способами.

## Prompt:

Применение RetroRAG в промптах для GPT ## Ключевые принципы RetroRAG

Исследование RetroRAG предлагает ретроактивный подход к обработке информации, который позволяет: - Пересматривать и корректировать цепочки рассуждений - Различать исходные и выводные доказательства - Оценивать релевантность и атрибуцию информации - Итеративно улучшать ответы

## Пример промпта, использующего принципы RetroRAG

[======] # Задание: Многоэтапный исследовательский анализ

## Контекст Мне нужен глубокий анализ [тема], включающий несколько взаимосвязанных аспектов. Используй подход RetroRAG для обработки информации.

## Инструкции 1. **Начальный анализ**: - Сформулируй первичные выводы на основе известных тебе данных - Четко разделяй факты (исходные доказательства) и логические выводы (выводные доказательства) - Отмечай степень уверенности в каждом утверждении

**Ретроактивная проверка**: Определи, какая дополнительная информация необходима для подтверждения выводов Укажи потенциальные пробелы в рассуждениях Сформулируй 2-3 конкретных вопроса для дальнейшего исследования

**Итоговый анализ**:

Пересмотри первоначальные выводы с учетом всей информации Отметь, какие первоначальные предположения подтвердились или были опровергнуты Представь итоговые выводы с указанием их надежности ## Формат ответа Структурируй ответ по этапам анализа, явно обозначая: - Исходные доказательства (что известно наверняка) - Выводные доказательства (логические заключения) - Области неопределенности (что требует дополнительной проверки) [======]

## Как этот промпт использует принципы RetroRAG

**Ретроактивность** — промпт требует пересмотра первоначальных выводов после получения дополнительной информации, что соответствует ключевому принципу RetroRAG

**Разделение доказательств** — явное разграничение между исходными фактами и логическими выводами помогает контролировать качество рассуждений

**Оценка релевантности** — промпт просит оценивать степень уверенности в утверждениях и выявлять пробелы в информации

**Итеративность** — структура промпта предполагает несколько этапов анализа с постепенным уточнением информации

**Ограниченное число итераций** — промпт содержит конкретное количество этапов (3), что соответствует рекомендации исследования об оптимальном числе итераций (3-5)

Такой подход значительно снижает риск галлюцинаций модели при работе со сложными многоэтапными задачами.