

Научиться задавать вопросы: Когда LLM-агенты сталкиваются с неясными инструкциями

Дата: 2025-02-16 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2409.00557>

Рейтинг: 75

Адаптивность: 85

Ключевые выводы:

Исследование направлено на изучение проблемы использования инструментов (API) языковыми моделями (LLM) при нечетких или неполных инструкциях пользователей. Основной результат - разработка нового метода Ask-when-Needed (AwN), который значительно улучшает способность LLM запрашивать уточнения у пользователей при неясных инструкциях, что повышает точность выполнения задач.

Объяснение метода:

Исследование предлагает ценную концепцию проактивного запроса уточнений при неясных инструкциях и классификацию типичных проблем, что помогает пользователям формулировать более эффективные запросы. Основные принципы могут быть легко адаптированы обычными пользователями в их промптах. Техническая реализация бенчмарка и автооценщика имеет ограниченную применимость для широкой аудитории.

Ключевые аспекты исследования: 1. **Проблема неясных инструкций:** Исследование выявляет критическую проблему - LLM-агенты часто сталкиваются с неясными инструкциями пользователей при использовании инструментов (API), что приводит к ошибкам выполнения.

Классификация проблем с инструкциями: Авторы проанализировали реальные пользовательские запросы и классифицировали проблемы на 4 категории: отсутствие ключевой информации (56%), множественные референции (11.3%), ошибки в инструкциях (17.3%) и запросы вне возможностей инструментов (15.3%).

Метод Ask-when-Needed (AwN): Предложен новый подход, при котором LLM-агенты проактивно задают уточняющие вопросы пользователю при обнаружении неясностей в запросе, вместо произвольной генерации недостающих аргументов.

Benchmark Noisy-ToolBench: Создан специализированный набор данных для оценки способности LLM распознавать неоднозначности в запросах пользователей и задавать уточняющие вопросы.

Автоматический оценщик ToolEvaluator: Разработан инструмент для автоматизации взаимодействия с LLM-агентами и оценки их производительности без участия человека.

Дополнение:

Применимость методов в стандартном чате

Методы исследования **не требуют дообучения или специального API** для основной концепции. Хотя авторы использовали специальные инструменты для экспериментов и оценки, ключевая идея Ask-when-Needed (AwN) может быть реализована в стандартном чате через промпт-инжиниринг.

Концепции, применимые в стандартном чате:

Проактивное запрашивание уточнений: Пользователи могут включать в промпты инструкции вроде "Если информации недостаточно для ответа, задай уточняющие вопросы вместо предположений".

Классификация типичных проблем: Пользователи могут проверять свои запросы на наличие четырех типов проблем перед отправкой (отсутствие ключевой информации, множественные референции, ошибки, запросы вне возможностей).

Структурированный диалог: Подход с пошаговым выполнением задач и проверкой наличия всей необходимой информации на каждом этапе.

Ожидаемые результаты:

- Снижение "галлюцинаций" и произвольных предположений LLM
- Более точные и релевантные ответы
- Повышение ответственности пользователя за качество запроса
- Формирование более эффективного диалога между пользователем и LLM

Анализ практической применимости: 1. **Проблема неясных инструкций** - Прямая применимость: Высокая. Исследование показывает, что пользователи должны осознавать важность четких инструкций при работе с LLM для получения точных результатов. - Концептуальная ценность: Очень высокая. Понимание того, что LLM могут "галлюцинировать" при неясных инструкциях вместо запроса дополнительной информации, критично для эффективного взаимодействия. - Потенциал для адаптации: Высокий. Пользователи могут научиться предоставлять более полные инструкции, предвидя возможные неясности.

Классификация проблем с инструкциями Прямая применимость: Средняя.

Пользователи могут использовать классификацию для самопроверки своих запросов. Концептуальная ценность: Высокая. Понимание типичных проблем помогает пользователям формулировать более эффективные запросы. Потенциал для адаптации: Высокий. Категории ошибок могут стать основой для чек-листа при составлении запросов.

Метод Ask-when-Needed (AwN)

Прямая применимость: Средняя. Пользователи могут адаптировать принцип, явно предлагая LLM задавать вопросы при неясностях. Концептуальная ценность: Высокая. Понимание важности диалога с LLM для уточнения запросов. Потенциал для адаптации: Высокий. Пользователи могут включать в свои промпты инструкции запрашивать уточнения при неясностях.

Benchmark Noisy-ToolBench

Прямая применимость: Низкая для обычных пользователей, высокая для разработчиков. Концептуальная ценность: Средняя. Демонстрирует типы сложных ситуаций. Потенциал для адаптации: Средний. Примеры из бенчмарка могут использоваться для обучения.

Автоматический оценщик ToolEvaluator

Прямая применимость: Низкая для обычных пользователей. Концептуальная ценность: Низкая для широкой аудитории. Потенциал для адаптации: Низкий для обычных пользователей.

Prompt:

Применение исследования Ask-when-Needed в промптах для GPT ## Ключевые инсайты исследования для промптинга

Исследование показывает, что языковые модели часто сталкиваются с неясными инструкциями пользователей, особенно когда требуется использование API или инструментов. Метод Ask-when-Needed (AwN) значительно улучшает способность моделей задавать уточняющие вопросы только когда это необходимо, что повышает точность выполнения задач.

Пример промпта с применением AwN

[=====] # Запрос на выполнение задачи с использованием метода Ask-when-Needed

Ты - ассистент, который помогает пользователям работать с API для [описание сервиса]. Следуй этому структурированному подходу:

Анализ полноты информации: Проверь, содержит ли запрос пользователя всю необходимую информацию для вызова API Определи, к какому типу проблем может

относиться запрос: Отсутствие ключевой информации Наличие множественных ссылок Инструкции с ошибками Запрос, выходящий за рамки возможностей инструмента

Проактивное уточнение (только при необходимости):

Если информации недостаточно, задай КОНКРЕТНЫЙ уточняющий вопрос Задавай вопросы только когда это действительно необходимо Формулируй вопросы четко, указывая какой именно параметр требуется уточнить

Выполнение задачи:

После получения всей необходимой информации, четко объясни какие действия будут выполнены Сформируй корректный вызов API с полученными параметрами Представь результат в понятной форме Начни с анализа моего запроса и действуй согласно методу Ask-when-Needed.

Мой запрос: [запрос пользователя] [=====]

Как работает данный промпт на основе исследования

Структурированный анализ: Промпт инструктирует модель сначала проанализировать полноту информации, что соответствует первому этапу метода AwN из исследования.

Классификация проблем: Включает типологию проблемных инструкций из исследования (56% - отсутствие ключевой информации, 11.3% - множественные ссылки и т.д.).

Уточнение по необходимости: Ключевой элемент AwN - запрашивать уточнения только когда это действительно необходимо, вместо автоматических вопросов или попыток угадать параметры.

Конкретные вопросы: Исследование показало, что конкретные уточняющие вопросы значительно повышают метрики A1 (правильность вопросов), A2 (точность вызова API) и A3 (качество ответов).

Структурированное выполнение: После получения всей информации модель выполняет задачу в соответствии с четкой структурой, что также повышает точность согласно исследованию.

Этот подход особенно эффективен для задач, где требуется взаимодействие с API или инструментами, и позволяет избежать как избыточных вопросов, так и ошибочных предположений.