

Рассуждения об аффордансах: причинное и композиторское рассуждение в LLM

Дата: 2025-02-23 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2502.16606>

Рейтинг: 78

Адаптивность: 85

Ключевые выводы:

Исследование направлено на оценку способностей больших языковых моделей (LLM) к причинно-следственному и композиционному мышлению в области использования объектов не по прямому назначению. Основные результаты показали, что современные модели (GPT-4o с цепочкой рассуждений и Claude 3.5) способны выполнять задачи на уровне, близком к человеческому, демонстрируя значительный прогресс по сравнению с более ранними моделями (GPT-3.5 и Claude 3).

Объяснение метода:

Исследование демонстрирует существенный прогресс в способности LLM к каузальному и композиционному мышлению. Ценность для пользователей включает: понимание различий между моделями, эффективность CoT-промптинга, и инсайты о том, как модели обрабатывают задачи, требующие творческого мышления. Методы непосредственно применимы в повседневном взаимодействии с LLM для получения более качественных ответов.

Ключевые аспекты исследования: 1. **Исследование каузального и композиционного мышления у LLM** - авторы проверяют способность языковых моделей выполнять задачи, требующие понимания причинно-следственных связей и композиционного мышления в области объектных возможностей.

Методология "инновации инструментов" - участникам (людям и LLM) предлагались задачи, где нужно выбрать нестандартный предмет для замены типичного инструмента (например, использовать кастрюлю вместо молотка).

Сравнительный анализ моделей разных поколений - исследование демонстрирует значительный прогресс между GPT-3.5/Claude 3 и GPT-4o/Claude 3.5 в способности выявлять и гибко применять причинно-значимые свойства объектов.

Chain-of-Thought (CoT) промптинг - исследование показывает, что CoT существенно улучшает производительность моделей, позволяя GPT-4o достигать

результатов на уровне человека.

Мультимодальное тестирование - во втором эксперименте авторы исследовали, как тип ввода (текст vs изображения) влияет на способность моделей решать задачи.

Дополнение:

Применимость методов в стандартном чате

Исследование не требует дообучения или API для применения основных методов. Ключевые подходы можно использовать в стандартном чате:

Chain-of-Thought (CoT) промптинг - исследование наглядно демонстрирует, что простое добавление инструкций "оцени каждый вариант отдельно, затем сделай выбор" значительно улучшает производительность моделей. Любой пользователь может применить эту технику в стандартном диалоге.

Структурированное рассуждение - качественный анализ показывает, что успешные модели (GPT-4o, Claude 3.5) последовательно выделяют причинно-значимые свойства объектов. Пользователи могут запрашивать LLM действовать подобным образом.

Композиционное мышление - понимание того, что LLM могут разбивать объекты на абстрактные свойства и комбинировать их по-новому, позволяет формулировать запросы, требующие креативного мышления.

Мультимодальный ввод - исследование показывает, что некоторые модели (GPT-4o) одинаково хорошо справляются с текстовыми и визуальными данными, что можно использовать в стандартных чатах с поддержкой изображений.

Применяя эти подходы, пользователи могут получать более качественные и творческие ответы от LLM, даже без специальных технических знаний или инструментов.

Анализ практической применимости: **Каузальное и композиционное мышление LLM**: - Прямая применимость: Пользователи могут использовать модели последнего поколения (GPT-4o, Claude 3.5) для решения задач, требующих выявления причинно-следственных связей между свойствами объектов. - Концептуальная ценность: Исследование демонстрирует, что современные LLM могут выходить за рамки чистых ассоциаций и применять гибкое мышление к новым ситуациям. - Потенциал для адаптации: Методы причинного мышления можно применять в различных контекстах — от решения бытовых проблем до профессиональных задач.

Методология "инновации инструментов": - Прямая применимость: Пользователи могут использовать подобные задачи для проверки способности LLM предлагать нестандартные решения. - Концептуальная ценность: Понимание того, как формулировать запросы, требующие от LLM креативного мышления, а не просто

ассоциаций. - Потенциал для адаптации: Этот подход можно использовать для различных областей, где требуется творческое решение проблем.

Сравнение моделей разных поколений: - Прямая применимость: Пользователи получают информацию о том, какие модели лучше справляются с задачами, требующими причинного мышления. - Концептуальная ценность: Демонстрация эволюции способностей LLM и понимание их прогресса. - Потенциал для адаптации: Знание о различиях между моделями помогает пользователям выбирать подходящий инструмент для конкретных задач.

Chain-of-Thought промптинг: - Прямая применимость: Пользователи могут применять CoT-промптинг для получения более качественных ответов от LLM. - Концептуальная ценность: Понимание того, как структурирование мышления модели влияет на качество результатов. - Потенциал для адаптации: Техника CoT может быть адаптирована для различных типов задач, требующих рассуждения.

Мультимодальное тестирование: - Прямая применимость: Информация о том, как различные модели справляются с текстовыми и визуальными входными данными. - Концептуальная ценность: Понимание ограничений моделей в работе с разными модальностями. - Потенциал для адаптации: Знание о различиях в обработке разных типов данных помогает пользователям формулировать запросы соответствующим образом.

Prompt:

Использование знаний из исследования об аффордансах в промптах для GPT ##
Ключевые выводы исследования для применения в промптах

Исследование показывает, что современные LLM способны к причинно-следственному и композиционному мышлению при правильном подходе к составлению промптов. Вот основные принципы, которые можно применить:

Использование цепочки рассуждений (CoT) значительно улучшает результаты **Ограничение количества вариантов** повышает точность ответов **Разбиение сложных задач** на подзадачи повышает эффективность **Явное указание на композиционное мышление** помогает моделям лучше выявлять абстрактные свойства объектов ## Пример промпта с применением этих принципов

[=====] # Задача: Найти нестандартное решение проблемы

Мне нужно закрепить плакат на стене, но у меня нет скотча или кнопок. У меня есть следующие предметы: - Зубная паста - Жевательная резинка - Мед - Шампунь

Инструкции: 1. Рассмотрите каждый предмет отдельно. 2. Для каждого предмета определите его ключевые физические свойства (липкость, вязкость и т.д.). 3. Оцените, как эти свойства могут помочь в решении моей задачи. 4. Выберите наиболее подходящий предмет и объясните, почему он лучше других. 5. Предложите конкретный

способ использования выбранного предмета для решения задачи.

Пожалуйста, сначала проведи подробный анализ каждого варианта, а потом сделай окончательный выбор. [=====]

Объяснение эффективности данного промпта

Применение CoT: промпт явно просит модель рассмотреть каждый вариант отдельно и провести пошаговый анализ перед принятием решения, что согласно исследованию повышает точность с ~40-50% до ~85%.

Ограниченное количество вариантов: используется только 4 варианта, что соответствует оптимальному количеству согласно исследованию (в условии Distractor с 9 вариантами производительность значительно падала).

Фокус на композиционном мышлении: промпт явно просит определить ключевые свойства объектов и оценить их применимость в новом контексте, что задействует способность модели к композиционному мышлению.

Структурированный подход: задача разбита на четкие подзадачи, что помогает модели последовательно применять причинно-следственное рассуждение.

Такой промпт позволяет максимально использовать способности современных LLM к нестандартному мышлению и применению объектов не по прямому назначению, что было продемонстрировано в исследовании.