Наличие личностей у ИИ приводит к болееHuman-like reasoning

Дата: 2025-02-21 00:00:00

Ссылка на исследование: https://arxiv.org/pdf/2502.14155

Рейтинг: 72
Адаптивность: 85

Ключевые выводы:

Основная цель исследования - изучить, могут ли большие языковые модели (LLM) эмулировать полный спектр человеческого мышления, включая как интуитивные (System 1), так и обдуманные (System 2) процессы рассуждения. Главный результат: LLM могут имитировать распределение человеческих ответов, особенно при использовании промптов с разными личностными характеристиками, причем открытые модели Llama и Mistral неожиданно превзошли проприетарные модели GPT.

Объяснение метода:

Исследование предлагает легко применимую технику персонализированного промптирования, позволяющую получать более человекоподобные и разнообразные ответы от LLM. Понимание различий между интуитивным и аналитическим мышлением помогает пользователям формулировать более эффективные запросы. Некоторые технические аспекты имеют ограниченную прямую применимость для широкой аудитории.

Ключевые аспекты исследования: 1. **Персонализированное промптирование для имитации человеческого мышления**: Исследование показывает, что придание LLM различных "личностных черт" через промпты на основе модели Big Five (открытость, добросовестность, экстраверсия, доброжелательность, нейротизм) позволяет им лучше имитировать разнообразие человеческих рассуждений.

Полный спектр рассуждений: Авторы вводят концепцию "проблемы полного спектра рассуждений" - задачу моделирования не только "правильных" ответов, но и всего спектра возможных человеческих рассуждений, включая интуитивные (Система 1) и аналитические (Система 2).

Расширенный формат NLI: Исследователи предлагают расширенную шестибалльную шкалу для задачи естественного языкового вывода (NLI), что позволяет более детально моделировать нюансы человеческих рассуждений по сравнению с традиционной трехбалльной шкалой.

Оптимизация с помощью генетического алгоритма: Для улучшения точности моделирования авторы применяют генетический алгоритм, оптимизирующий веса различных "личностных" промптов, что значительно улучшает способность LLM предсказывать распределение человеческих ответов.

Превосходство моделей с открытым исходным кодом: Исследование обнаружило, что открытые модели (Llama, Mistral) превосходят закрытые модели (GPT) в задаче имитации человеческих рассуждений, что противоречит распространенному мнению о превосходстве закрытых моделей.

Дополнение: Исследование не требует дообучения моделей или специального API для применения основных концепций. Основные методы и подходы могут быть реализованы в стандартном чате с любой LLM.

Ключевые концепции, которые можно применить в стандартном чате:

Персонализированные промпты на основе личностных черт: Пользователи могут включать в запросы инструкции типа "Отвечай как человек с высокой открытостью к опыту и творческим мышлением" или "Отвечай как консервативный, осторожный человек". Это позволит получать более разнообразные ответы, отражающие различные стили мышления.

Явное указание на тип мышления: Пользователи могут запрашивать либо быстрые интуитивные ответы (Система 1), либо тщательно обдуманные аналитические рассуждения (Система 2) с помощью инструкций вроде "Дай мне быстрый, интуитивный ответ" или "Подумай тщательно и шаг за шагом".

Комбинирование различных "личностей": Вместо использования генетического алгоритма пользователи могут последовательно запрашивать ответы с разными личностными чертами, а затем синтезировать из них наиболее полезные элементы.

Расширенная шкала уверенности: Можно адаптировать шестибалльную шкалу для получения более нюансированных ответов, например: "Оцени вероятность этого утверждения по шкале от 1 до 6, где 1 - абсолютно ложно, а 6 - абсолютно истинно".

Ожидаемые результаты от применения этих концепций: - Более разнообразные и творческие ответы - Возможность получать как быстрые интуитивные, так и глубокие аналитические рассуждения - Более нюансированные оценки вероятности и уверенности - Улучшенное понимание разных точек зрения на одну и ту же проблему

Хотя исследователи использовали специальные методы для валидации своего подхода (например, генетические алгоритмы), сами основные концепции не требуют технической реализации и могут быть применены в обычном диалоге с LLM.

Анализ практической применимости: **1. Персонализированное промптирование для имитации человеческого мышления** - **Прямая применимость**: Высокая.

Пользователи могут немедленно применять промпты с личностными характеристиками для получения более разнообразных и человекоподобных ответов от LLM, что особенно полезно при творческой работе или генерации контента. - **Концептуальная ценность**: Значительная. Понимание того, что LLM можно "настраивать" через личностные промпты, дает пользователям более глубокое понимание возможностей и ограничений моделей. - **Потенциал для адаптации**: Очень высокий. Техника персонализированного промптирования может быть легко адаптирована к различным задачам и контекстам.

2. Полный спектр рассуждений - **Прямая применимость**: Средняя. Обычным пользователям может быть сложно напрямую применять эту концепцию, но понимание того, что LLM могут моделировать различные типы рассуждений, полезно. - **Концептуальная ценность**: Высокая. Понимание различий между интуитивным и аналитическим мышлением помогает пользователям формулировать более эффективные запросы. - **Потенциал для адаптации**: Высокий. Пользователи могут адаптировать свои промпты для получения либо быстрых интуитивных ответов, либо более обдуманных аналитических рассуждений.

3. Расширенный формат NLI - **Прямая применимость**: Низкая для обычных пользователей, поскольку требует понимания технических аспектов NLI. - **Концептуальная ценность**: Средняя. Показывает, как можно улучшить детализацию ответов LLM, что может быть полезно в определенных контекстах. - **Потенциал для адаптации**: Средний. Концепция расширенных шкал для ответов может быть адаптирована для других задач, требующих нюансированных ответов.

4. Оптимизация с помощью генетического алгоритма - **Прямая применимость**: Низкая для обычных пользователей из-за технической сложности. - **Концептуальная ценность**: Средняя. Демонстрирует важность оптимизации весов различных промптов. - **Потенциал для адаптации**: Средний. Принцип комбинирования различных промптов может быть упрощен и адаптирован обычными пользователями без применения генетических алгоритмов.

5. Превосходство моделей с открытым исходным кодом - **Прямая применимость**: Средняя. Пользователи могут предпочесть использование открытых моделей для задач, требующих человекоподобных рассуждений. - **Концептуальная ценность**: Высокая. Развенчивает миф о том, что только самые крупные проприетарные модели способны к человекоподобным рассуждениям. - **Потенциал для адаптации**: Высокий. Пользователи могут экспериментировать с различными моделями для разных задач, основываясь на этих выводах.

Сводная оценка полезности: Предварительная оценка полезности: 75

Исследование предлагает высоко применимую технику персонализированного промптирования, которая может быть немедленно использована широкой аудиторией для получения более разнообразных и человекоподобных ответов от LLM. Концепция использования личностных черт в промптах интуитивно понятна и не требует специальных технических знаний для применения.

Особенно ценным является понимание того, что LLM могут моделировать как интуитивное (Система 1), так и аналитическое (Система 2) мышление, что позволяет пользователям более осознанно формулировать запросы в зависимости от желаемого типа ответа.

Контраргументы, почему оценка могла бы быть выше: 1. Техника персонализированного промптирования исключительно проста в применении и может быть использована любым пользователем без технических навыков. 2. Открытие того, что открытые модели превосходят закрытые в имитации человеческих рассуждений, имеет широкую практическую ценность.

Контраргументы, почему оценка могла бы быть ниже: 1. Некоторые аспекты исследования, такие как шестибалльная шкала NLI и генетические алгоритмы, имеют ограниченную прямую применимость для обычных пользователей. 2. Исследование сосредоточено на академическом аспекте моделирования человеческих рассуждений, а не на практических применениях этой техники.

После рассмотрения этих аргументов, корректирую оценку до 72. Хотя исследование предлагает высоко применимую технику персонализированного промптирования, некоторые аспекты слишком академические для широкого применения.

Оценка полезности: 72

Эта оценка обоснована следующими факторами: 1. Техника персонализированного промптирования легко применима и не требует технических навыков. 2. Исследование предлагает концептуально важное понимание возможностей LLM в моделировании различных типов человеческого мышления. 3. Выводы о превосходстве открытых моделей имеют практическую ценность. 4. Некоторые аспекты исследования (шестибалльная шкала NLI, генетические алгоритмы) имеют ограниченную прямую ценность для широкой аудитории. 5. Исследование в большей степени академическое, чем практическое, но его основные выводы могут быть легко адаптированы.

Уверенность в оценке: Очень сильная. Я тщательно проанализировал основные аспекты исследования и их применимость для широкой аудитории. Техника персонализированного промптирования представляет собой непосредственно применимый метод, который любой пользователь может начать использовать сразу же, а концептуальное разделение на интуитивное и аналитическое мышление дает важное понимание работы LLM. Эти ключевые аспекты имеют высокую практическую ценность, в то время как более технические аспекты исследования имеют ограниченную прямую применимость, что обосновывает итоговую оценку.

Оценка адаптивности: Адаптивность: 85

Исследование демонстрирует высокий потенциал для адаптации по следующим причинам:

Техника персонализированного промптирования с использованием личностных черт легко адаптируется к стандартным чатам без необходимости в дополнительных API или дообучении. Пользователи могут просто включать в свои запросы фразы, определяющие личностные черты, чтобы получать более разнообразные ответы.

Концепция моделирования как интуитивного (Система 1), так и аналитического (Система 2) мышления может быть легко адаптирована путем включения соответствующих инструкций в промпты (например, "ответь быстро, интуитивно" против "обдумай этот вопрос тщательно").

Хотя генетический алгоритм для оптимизации весов промптов технически сложен, сама идея комбинирования различных промптов может быть адаптирована обычными пользователями через более простые методы.

Выводы о превосходстве открытых моделей в задачах имитации человеческих рассуждений могут направлять выбор пользователей при работе с различными LLM.

Метод не требует технического доступа к архитектуре модели и может быть реализован в обычном чате, что делает его исключительно адаптивным для широкой аудитории.

|| <Оценка: 72> || <Объяснение: Исследование предлагает легко применимую технику персонализированного промптирования, позволяющую получать более человекоподобные и разнообразные ответы от LLM. Понимание различий между интуитивным и аналитическим мышлением помогает пользователям формулировать более эффективные запросы. Некоторые технические аспекты имеют ограниченную прямую применимость для широкой аудитории.> || <Адаптивность: 85>

Prompt:

Использование знаний из исследования о личностях ИИ в промптах для GPT
Ключевые выводы исследования для промптинга

Исследование показало, что использование промптов с различными личностными характеристиками (personality prompting) значительно улучшает способность языковых моделей имитировать разнообразие человеческого мышления, включая как интуитивные (System 1), так и обдуманные (System 2) процессы.

Пример промпта с использованием личностных характеристик

[======] Я хочу, чтобы ты выступил в роли консультанта по маркетингу с определенными личностными характеристиками.

Твой профиль: Ты очень открыт к новому опыту, креативен и любознателен. При этом ты достаточно организован и ответственен, но не слишком консервативен в

своих взглядах.

Задача: Проанализируй предложенную маркетинговую стратегию для нового продукта на рынке смартфонов и предложи 3-4 нестандартных идеи, которые могли бы выделить продукт среди конкурентов.

Пожалуйста, сначала дай свою быструю интуитивную реакцию (System 1), а затем более обдуманный аналитический ответ (System 2). Для оценки каждой идеи используй 6-балльную шкалу потенциальной эффективности от "крайне неэффективно" до "крайне эффективно".

Маркетинговая стратегия: [описание стратегии] [======]

Как работают знания из исследования в этом промпте

Использование личностных характеристик - промпт задает конкретный личностный профиль (открытость к опыту, креативность, организованность), что согласно исследованию помогает получить более разнообразные и человекоподобные рассуждения.

Разделение на System 1 и System 2 - промпт явно запрашивает как быструю интуитивную реакцию, так и медленное аналитическое мышление, что отражает двойную систему человеческого мышления, исследованную в работе.

6-вариантная шкала - вместо стандартной 3-балльной шкалы используется 6-балльная, что, согласно исследованию, позволяет получить более детальное и близкое к человеческому распределение оценок.

Сочетание творческого и аналитического подходов - промпт балансирует между открытостью к новому (для генерации креативных идей) и организованностью (для их структурированного анализа), что отражает оптимизированные комбинации личностных черт из исследования.

Применяя эти принципы, вы можете создавать промпты, которые будут вызывать более разнообразные, естественные и человекоподобные ответы от GPT для различных задач.