

# MultiAgentBench: Оценка сотрудничества и конкуренции многопользовательских агентов

Дата: 2025-03-03 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2503.01935>

Рейтинг: 60

Адаптивность: 75

## Ключевые выводы:

Исследование представляет Multi-Agent Bench - комплексный бенчмарк для оценки систем на основе LLM, работающих в многоагентном режиме. Основная цель - оценить не только выполнение задач, но и качество сотрудничества и конкуренции между агентами. Главные результаты показывают, что GPT-4o-mini достигает наивысших показателей выполнения задач, графовая структура координации показывает лучшие результаты в исследовательских сценариях, а когнитивное планирование улучшает достижение ключевых этапов на 3%.

## Объяснение метода:

Исследование представляет ценные концепции мультиагентной координации и протоколы взаимодействия, полезные для разработчиков. Концептуально демонстрирует эффективность разных топологий взаимодействия и стратегий планирования. Однако требует значительной технической адаптации для применения обычными пользователями и специализированной инфраструктуры для полной реализации.

Ключевые аспекты исследования: 1. **MultiAgentBench** - Комплексный бенчмарк для оценки систем на основе LLM-агентов в разнообразных сценариях сотрудничества и конкуренции. Включает шесть различных интерактивных сценариев, от исследовательских задач до игр и переговоров.

**Оценка координации и коммуникации** - Исследование вводит новые метрики для оценки не только успешности выполнения задач, но и качества сотрудничества между агентами, включая KPI на основе достижения этапов, оценки планирования и коммуникации.

**Протоколы координации** - Изучение различных топологий координации (звезда, цепочка, дерево, граф) и стратегий планирования (обычный промпт, цепочка мыслей, групповое обсуждение, когнитивное планирование).

**Эмерджентные социальные поведения** - Исследование выявляет возникающие

социальные поведения у LLM-агентов, такие как стратегический обмен информацией, поляризованное сотрудничество и адаптация стратегий на основе ролей.

**Инфраструктура MARBLE** - Предлагается фреймворк для мультиагентной координации, включающий координационные механизмы, когнитивные модули и инструменты взаимодействия с окружающей средой.

Дополнение:

## **# Применимость в стандартном чате без дообучения или API**

Исследование действительно использует API и специализированную инфраструктуру, однако ключевые концепции и подходы можно адаптировать для стандартного чата без необходимости дообучения моделей. Авторы использовали расширенные техники больше для удобства исследования и систематизации результатов.

## **# Концепции и подходы для стандартного чата:**

**Протоколы координации** - Можно реализовать различные топологии взаимодействия (звезда, цепочка, дерево, граф) через правильное структурирование промптов и ролей в обычном чате. Например: **Звезда**: Один центральный агент (планировщик) координирует других специализированных агентов **Цепочка**: Последовательная передача результатов между агентами в определенном порядке **Граф**: Гибкая структура, где любой агент может взаимодействовать с другими по мере необходимости

**Стратегии планирования** - Когнитивное планирование показало лучшие результаты и может быть реализовано через:

Явное планирование задач перед их выполнением Итеративную проверку ожидаемых результатов против фактических Корректировку планов на основе полученного опыта

**Разделение на роли планировщиков и исполнителей** - Эффективный подход, который можно реализовать в стандартном чате:

Планировщик разбивает задачу на подзадачи и распределяет их Исполнители решают конкретные подзадачи Планировщик интегрирует результаты и корректирует план

**Групповое обсуждение** - Можно имитировать через:

Последовательное представление перспектив разных агентов по одной проблеме Синтез этих перспектив в единое решение

## # Ожидаемые результаты от применения:

**Повышение эффективности решения сложных задач** - Разделение задачи между агентами с разной специализацией улучшает качество решения **Более структурированные решения** - Четкое планирование и координация приводят к более логичным и последовательным результатам **Преодоление ограничений контекста** - Правильная координация позволяет эффективнее использовать ограниченный контекст модели **Улучшенное обнаружение ошибок** - Разные агенты могут проверять работу друг друга Эти подходы можно реализовать в стандартном чате через правильное структурирование промптов, без необходимости дополнительного API или дообучения моделей.

Анализ практической применимости: 1. **MultiAgentBench и разнообразные сценарии** - **Прямая применимость**: Средняя. Обычные пользователи не могут напрямую использовать этот бенчмарк, но разработчики систем могут применять его для тестирования. - **Концептуальная ценность**: Высокая. Демонстрирует разнообразие задач, где мультиагентный подход эффективен (исследования, кодирование, переговоры). - **Потенциал для адаптации**: Высокий. Сценарии можно адаптировать для создания специализированных мультиагентных систем для конкретных задач.

**Метрики оценки координации и коммуникации** **Прямая применимость**: Низкая для конечных пользователей, но высокая для разработчиков систем. **Концептуальная ценность**: Высокая. Понимание важности не только результата, но и процесса координации между агентами. **Потенциал для адаптации**: Средний. Метрики можно упростить для использования в пользовательских системах.

## Протоколы координации и стратегии планирования

**Прямая применимость**: Средняя. Пользователи могут адаптировать идею различных топологий для своих систем. **Концептуальная ценность**: Очень высокая. Понимание сильных и слабых сторон разных стратегий координации. **Потенциал для адаптации**: Высокий. Принципы можно применить для улучшения взаимодействия нескольких агентов в чате.

## Эмерджентные социальные поведения

**Прямая применимость**: Низкая. Требуется сложная система для наблюдения таких поведений. **Концептуальная ценность**: Высокая. Понимание, что LLM-агенты могут развивать сложные социальные динамики. **Потенциал для адаптации**: Средний. Знания о социальных динамиках могут помочь в проектировании более естественных мультиагентных систем.

## Инфраструктура MARBLE

**Прямая применимость**: Низкая для обычных пользователей, высокая для разработчиков. **Концептуальная ценность**: Высокая. Предоставляет структуру для

создания мультиагентных систем. **Потенциал для адаптации:** Высокий. Элементы фреймворка могут быть адаптированы для различных приложений. Сводная оценка полезности: Предварительная оценка: 65/100

Исследование предоставляет ценные концепции и методологии для создания и оценки мультиагентных систем на основе LLM. Оно особенно полезно для разработчиков и исследователей, но имеет ограниченную прямую применимость для обычных пользователей LLM.

Контраргументы для более высокой оценки: 1. Исследование демонстрирует конкретные протоколы координации и стратегии планирования, которые могут быть адаптированы даже обычными пользователями для создания более эффективных систем. 2. Понимание эмерджентных социальных поведений может значительно улучшить взаимодействие пользователей с мультиагентными системами.

Контраргументы для более низкой оценки: 1. Большая часть исследования требует технических знаний и доступа к API моделей, что недоступно для большинства пользователей. 2. Воспроизведение результатов требует значительных вычислительных ресурсов и специализированной инфраструктуры.

После рассмотрения контраргументов, скорректированная оценка: 60/100

Основания для оценки: 1. Высокая ценность для разработчиков и исследователей мультиагентных систем. 2. Хорошая концептуальная ценность в понимании принципов координации. 3. Ограниченная прямая применимость для обычных пользователей LLM. 4. Необходимость значительной адаптации для практического применения в широкодоступных системах. 5. Высокий потенциал для будущего развития пользовательских мультиагентных систем.

Уверенность в оценке: Уверенность: очень сильная.

Исследование подробно описывает методологию, результаты и практические аспекты мультиагентных систем. Оценка учитывает как техническую глубину работы, так и потенциальную полезность для различных категорий пользователей LLM. Также учтены конкретные примеры применения и ограничения, что повышает точность оценки.

Оценка адаптивности: Оценка адаптивности: 75/100

Принципы и концепции исследования имеют высокий потенциал для адаптации в обычном чате. Идеи различных топологий координации (звезда, цепочка, дерево, граф) могут быть применены для структурирования взаимодействия между агентами даже в базовом чате.

Пользователи могут извлечь ценные идеи о стратегиях планирования, особенно когнитивное планирование и групповое обсуждение, которые показали эффективность в улучшении координации.

Высокий потенциал для внедрения выводов исследования в будущее взаимодействия с LLM. Понимание того, как агенты обмениваются информацией и адаптируют свои стратегии на основе ролей, может значительно улучшить проектирование мультиагентных систем.

Специализированные методы, такие как KPI на основе достижения этапов и оценка коммуникации, могут быть абстрагированы до общих принципов для оценки эффективности взаимодействия в любой мультиагентной системе.

Исследование предлагает концепции, которые можно применить даже без доступа к специализированной инфраструктуре, что повышает его адаптивность для широкого круга сценариев использования.

|| <Оценка: 60> || <Объяснение: Исследование представляет ценные концепции мультиагентной координации и протоколы взаимодействия, полезные для разработчиков. Концептуально демонстрирует эффективность разных топологий взаимодействия и стратегий планирования. Однако требует значительной технической адаптации для применения обычными пользователями и специализированной инфраструктуры для полной реализации.> || <Адаптивность: 75>

## **Prompt:**

Использование знаний из исследования MultiAgentBench в промптах для GPT  
**Ключевые выводы исследования для промптов**

Исследование MultiAgentBench предоставляет ценную информацию о том, как оптимально выстраивать взаимодействие между несколькими LLM-агентами. Эти знания можно эффективно применить при создании промптов для GPT, особенно когда требуется решение сложных задач с использованием нескольких "агентов" в рамках одной сессии.

## **Пример промпта на основе исследования**

[=====]

Многоагентное исследование рынка  
**Структура и роли**

Ты будешь действовать как система из 3 агентов с графовой структурой координации: 1. Аналитик данных - собирает и анализирует информацию о рынке 2. Маркетолог - интерпретирует данные с точки зрения потребительского поведения 3. Стратег - формулирует итоговые рекомендации

## Метод координации

- Используй графовую структуру взаимодействия, где каждый агент может напрямую общаться с любым другим
- Применяй когнитивное самоэволюционирующее планирование: в начале работы каждого агента формулируй ожидаемый результат, а после выполнения сравнивай его с фактическим
- Ограничь обсуждение 5-7 итерациями для оптимальной эффективности

## Процесс работы

Сначала представь план исследования для каждого агента с ожидаемыми результатами. Проведи 5 итераций обсуждения, где агенты обмениваются информацией. После каждой итерации проводи самооценку и корректировку планов. В финальном отчете представь консолидированные выводы и рекомендации.

### Задача

Исследуй рынок электросамокатов в городской среде и разработай стратегию вывода нового продукта. [=====]

## Объяснение применения знаний из исследования

**Графовая структура координации** - исследование показало, что графовый протокол превосходит другие (звезда, цепь, дерево) в исследовательских сценариях, поэтому промпт предусматривает возможность прямого взаимодействия между всеми агентами.

**Оптимальное количество агентов** - согласно исследованию, использование 3 агентов обеспечивает значительное улучшение координации без избыточной сложности.

**Когнитивное самоэволюционирующее планирование** - этот подход показал превосходную координацию, поэтому промпт включает формулировку ожидаемых результатов и последующее сравнение с фактическими.

**Ограничение итераций** - исследование показало, что оптимальное число итераций для сложных задач составляет около 7, после чего эффективность снижается, поэтому в промпте указано ограничение в 5-7 итераций.

**Стратегический обмен информацией** - промпт предусматривает структурированный обмен информацией между агентами с учетом их ролей и компетенций.

Такой подход к промптам позволяет максимально использовать возможности GPT для решения сложных задач, требующих многостороннего анализа и координации между различными перспективами.