

# Сравнительный анализ на основе DeepSeek, ChatGPT и Google Gemini: характеристики, техники, производительность, перспективы будущего.

Дата: 2025-02-25 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2503.04783>

Рейтинг: 72

Адаптивность: 80

## Ключевые выводы:

Исследование направлено на сравнительный анализ трех ведущих языковых моделей - DeepSeek, ChatGPT и Google Gemini. Основные результаты показывают, что каждая модель имеет свои уникальные преимущества: DeepSeek эффективен в узкоспециализированных задачах благодаря архитектуре Mixture of Experts (MoE), ChatGPT превосходит в разговорных задачах благодаря RLHF, а Gemini выделяется мультимодальными возможностями.

## Объяснение метода:

Исследование предоставляет ценное сравнение трех популярных LLM с подробными бенчмарками и анализом их архитектур, что позволяет пользователям делать обоснованный выбор модели для конкретных задач. Хотя исследование содержит значительный объем технической информации, понимание сильных и слабых сторон моделей напрямую применимо в повседневном использовании LLM.

Ключевые аспекты исследования: 1. **Сравнительный анализ архитектур:** Исследование детально сравнивает архитектурные особенности трех ведущих моделей - DeepSeek (использует Mixture of Experts), ChatGPT (использует плотную трансформерную модель с RLHF) и Google Gemini (использует мультимодальную трансформерную архитектуру).

**Анализ производительности:** Работа представляет подробные бенчмарки и сравнительные тесты по различным метрикам, включая рассуждение, знания, научное мышление, количественные рассуждения, кодирование и многоязычность.

**Данные для обучения:** Исследование анализирует наборы данных, используемые для обучения каждой модели, их состав и влияние на производительность моделей в различных задачах.

**Сильные и слабые стороны:** Авторы выявляют специфические преимущества каждой модели - эффективность DeepSeek для специализированных задач, быстрый разговорный отклик ChatGPT и мультимодальные возможности Gemini.

**Будущие направления развития:** Исследование предлагает обзор текущих проблем (включая баланс производительности с вычислительной эффективностью) и потенциальных направлений развития LLM.

Дополнение:

## **# Применимость методов в стандартном чате без дообучения/API**

Исследование не требует дообучения моделей или использования API для применения большинства его выводов. Основные концепции и подходы могут быть использованы непосредственно в стандартном чате:

**Выбор подходящей модели** - пользователи могут выбрать наиболее подходящую модель для своих задач: DeepSeek для узкоспециализированных задач (медицина, право, финансы) ChatGPT для разговорного взаимодействия и общих задач Gemini для мультимодальных задач (работа с текстом, изображениями, кодом)

**Адаптация запросов под сильные стороны модели:**

Для DeepSeek: формулирование специализированных, профессиональных запросов в конкретной области Для ChatGPT: использование разговорного стиля, многоступенчатых запросов Для Gemini: формулирование запросов с использованием различных модальностей (текст + изображения)

**Применение знаний о производительности в разных задачах:**

Для задач рассуждения и логики: предпочтение ChatGPT и DeepSeek Для кодирования: выбор между DeepSeek и ChatGPT в зависимости от сложности Для мультязычных задач: учет относительной производительности каждой модели

**Понимание ограничений моделей:**

Учет возможных галлюцинаций и предвзятостей в ответах Реалистичные ожидания от моделей в зависимости от их архитектуры и данных обучения Результаты от применения этих концепций: - Более эффективное использование LLM для конкретных задач - Улучшение качества получаемых ответов - Снижение разочарования от нереалистичных ожиданий - Экономия времени за счет выбора наиболее подходящей модели для конкретной задачи

Анализ практической применимости: **Сравнительный анализ архитектур:** - Прямая применимость: Средняя. Пользователи могут выбрать наиболее подходящую модель для своих задач (DeepSeek для узкоспециализированных задач, ChatGPT

для разговорного взаимодействия, Gemini для мультимодальных задач). - Концептуальная ценность: Высокая. Пользователи получают понимание различий между моделями и как эти различия влияют на их производительность. - Потенциал для адаптации: Средний. Знание архитектурных особенностей может помочь в формулировании более эффективных запросов к конкретным моделям.

**Анализ производительности:** - Прямая применимость: Высокая. Пользователи могут выбрать наиболее эффективную модель для конкретных задач на основе представленных бенчмарков. - Концептуальная ценность: Высокая. Четкое понимание сильных и слабых сторон каждой модели в различных задачах. - Потенциал для адаптации: Высокий. Знание о производительности в разных задачах помогает пользователям адаптировать свои запросы и ожидания.

**Данные для обучения:** - Прямая применимость: Низкая. Обычный пользователь не может изменить данные обучения. - Концептуальная ценность: Средняя. Понимание источников данных помогает осознать возможные ограничения и предвзятости моделей. - Потенциал для адаптации: Низкий. Сложно использовать эту информацию для адаптации запросов.

**Сильные и слабые стороны:** - Прямая применимость: Высокая. Пользователи могут выбрать модель, которая лучше всего подходит для их конкретных задач. - Концептуальная ценность: Высокая. Понимание ограничений помогает формировать реалистичные ожидания. - Потенциал для адаптации: Высокий. Знание о сильных и слабых сторонах помогает формулировать более эффективные запросы.

**Будущие направления развития:** - Прямая применимость: Низкая. Информация о будущих направлениях имеет ограниченную немедленную применимость. - Концептуальная ценность: Средняя. Понимание тенденций развития может помочь в стратегическом планировании использования LLM. - Потенциал для адаптации: Низкий. Сложно адаптировать текущие запросы на основе будущих возможностей.

Сводная оценка полезности: На основе проведенного анализа, предварительная оценка полезности исследования для широкой аудитории составляет 75 из 100.

Аргументы в пользу более высокой оценки: 1. Исследование предоставляет четкое сравнение трех популярных моделей, что напрямую помогает пользователям выбрать наиболее подходящую для их задач. 2. Подробные бенчмарки дают конкретные критерии для выбора модели в зависимости от задачи.

Аргументы в пользу более низкой оценки: 1. Значительная часть исследования посвящена техническим деталям архитектуры и обучения, которые имеют ограниченную практическую ценность для обычных пользователей. 2. Отсутствие конкретных рекомендаций по формулированию запросов для каждой модели снижает прямую применимость.

После рассмотрения этих аргументов, я корректирую оценку до 72 из 100. Исследование предоставляет ценную информацию для выбора подходящей модели и понимания её ограничений, но содержит значительный объем технической

информации, которая менее полезна для широкой аудитории.

Оценка дана по следующим причинам: 1. Исследование предоставляет четкое сравнение сильных и слабых сторон популярных LLM, что помогает пользователям делать обоснованный выбор. 2. Бенчмарки по различным задачам дают конкретные критерии для выбора модели. 3. Понимание архитектурных различий помогает пользователям формировать реалистичные ожидания. 4. Часть технической информации имеет ограниченную практическую ценность для широкой аудитории. 5. Отсутствуют конкретные рекомендации по оптимизации запросов для каждой модели.

Уверенность в оценке: Очень сильная. Исследование представляет собой комплексный сравнительный анализ с четкими метриками производительности и подробным описанием особенностей каждой модели. Эта информация напрямую полезна для выбора подходящей модели и понимания её возможностей и ограничений, что является ключевым аспектом для пользователей LLM.

Оценка адаптивности: Оценка адаптивности: 80 из 100.

Исследование предоставляет принципы и концепции, которые могут быть легко адаптированы для использования в обычном чате:

Пользователи могут выбрать наиболее подходящую модель для своих задач на основе представленных сравнений (DeepSeek для специализированных задач, ChatGPT для разговорного взаимодействия, Gemini для мультимодальных задач).

Понимание сильных и слабых сторон каждой модели позволяет пользователям формулировать более эффективные запросы и иметь реалистичные ожидания от результатов.

Знание о производительности моделей в различных задачах (рассуждение, знания, программирование) помогает пользователям адаптировать свои запросы для достижения лучших результатов.

Исследование предоставляет принципы, которые могут быть применены для оценки и других LLM, не рассмотренных в работе.

Информация о типах данных, используемых для обучения каждой модели, помогает пользователям понять возможные ограничения и предвзятости в ответах.

|| <Оценка: 72> || <Объяснение: Исследование предоставляет ценное сравнение трех популярных LLM с подробными бенчмарками и анализом их архитектур, что позволяет пользователям делать обоснованный выбор модели для конкретных задач. Хотя исследование содержит значительный объем технической информации, понимание сильных и слабых сторон моделей напрямую применимо в повседневном использовании LLM.> || <Адаптивность: 80>

## Prompt:

Использование знаний из исследования LLM в промптах

### Ключевые инсайты для применения в промптах

Исследование о DeepSeek, ChatGPT и Gemini предоставляет ценную информацию о сильных сторонах каждой модели, которую можно использовать для создания более эффективных промптов.

### Пример промпта с учетом результатов исследования

[=====] Я работаю над [узкоспециализированной финансовой задачей] и использую ChatGPT. Учитывая, что: 1. ChatGPT сильна в разговорных задачах благодаря RLHF 2. Техника chain-of-thought повышает точность рассуждений 3. Модель хорошо сохраняет контекст в длительных диалогах

Помоги мне проанализировать следующие финансовые данные, используя пошаговое рассуждение. Разбей анализ на четкие этапы, объясняя каждый шаг твоего рассуждения:

[финансовые данные]

После анализа, суммируй ключевые выводы и предложи три возможных стратегии действий, основанных на этих данных. [=====]

### Объяснение эффективности промпта

Данный промпт эффективен, поскольку:

**Использует сильные стороны конкретной модели** - учитывает, что ChatGPT хорошо справляется с разговорными задачами и сохранением контекста

**Применяет технику chain-of-thought** - исследование показало, что пошаговое рассуждение значительно улучшает производительность моделей в сложных задачах

**Структурирует запрос** - четко определяет задачу и ожидаемый формат ответа, что помогает модели сфокусироваться на релевантной информации

**Учитывает специфику задачи** - для финансовой области важна точность и последовательность рассуждения, что соответствует возможностям ChatGPT

### Как адаптировать промпты для разных моделей

- Для DeepSeek: Фокусируйтесь на узкоспециализированных технических задачах и

эффективном использовании ресурсов

- Для ChatGPT: Используйте диалоговый формат и техники улучшения рассуждений
- Для Gemini: Включайте мультимодальные элементы (текст + изображения) для комплексных задач

Исследование подчеркивает важность выбора правильной модели и техники промптинга для конкретной задачи, что может значительно повысить качество результатов.