

# Размышление в спектре: Согласование больших языковых моделей с мышлением Системы 1 и Системы 2

Дата: 2025-02-17 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2502.12470>

Рейтинг: 82

Адаптивность: 85

## Ключевые выводы:

Исследование направлено на изучение способности больших языковых моделей (LLM) адаптироваться между интуитивным (система 1) и аналитическим (система 2) стилями мышления. Основные результаты показывают, что модели, настроенные на систему 2, превосходят в арифметических и символических задачах, а модели системы 1 лучше справляются с задачами здравого смысла, демонстрируя компромисс между точностью и эффективностью.

## Объяснение метода:

Исследование предлагает практичную концепцию двух систем мышления (быстрой интуитивной и медленной аналитической), которую пользователи могут немедленно применять через промптинг. Результаты дают четкие рекомендации: использовать Систему 1 для задач здравого смысла и Систему 2 для математических задач. Концепция доступна для понимания широкой аудиторией и не требует технических знаний для применения.

## Ключевые аспекты исследования: 1. **Разделение режимов мышления:** Исследование рассматривает два типа мышления в LLM: быстрое интуитивное (Система 1) и медленное аналитическое (Система 2), основываясь на теориях когнитивной психологии.

**Создание специального датасета:** Авторы создали датасет из 2000 примеров, где каждый вопрос имеет два валидных ответа — быстрый интуитивный (Система 1) и подробный аналитический (Система 2).

**Выравнивание (alignment) моделей:** Исследователи обучили модели предпочитать либо быстрый интуитивный стиль мышления (Система 1), либо медленный аналитический (Система 2).

**Анализ эффективности различных режимов мышления:** Модели,

ориентированные на Систему 2, превосходят в арифметических и символических задачах, а модели Системы 1 лучше справляются с задачами здравого смысла.

**Изучение неопределенности в ответах:** Модели Системы 1 дают более уверенные и однозначные ответы, а модели Системы 2 проявляют большую неопределенность при рассуждениях.

## Дополнение:

### Применимость в стандартном чате без дообучения

Исследование демонстрирует, что для полноценной реализации подхода с разделением на Систему 1 и Систему 2 было использовано дообучение моделей. Однако **ключевые концепции и подходы можно эффективно применять в стандартном чате без дообучения** через промптинг.

#### Концепции для применения в стандартном чате:

**Выбор режима мышления через промпты:** Для задач, требующих пошагового анализа (математика, логика): "Рассуждай шаг за шагом, анализируя каждый аспект проблемы" (Система 2) Для задач здравого смысла: "Дай быстрый и интуитивный ответ на основе общих знаний" (Система 1)

**Адаптация под тип задачи:**

Исследование показало, что Система 2 эффективнее для арифметических и символических задач Система 1 лучше для задач здравого смысла Эту информацию можно использовать для выбора подходящего типа запроса

**Управление уверенностью в ответах:**

Если нужен однозначный ответ: "Дай прямой и однозначный ответ" (Система 1) Если важно видеть неопределенность: "Рассмотри разные варианты и их вероятность" (Система 2)

**Баланс между эффективностью и точностью:**

Для быстрых решений: "Ответь кратко, опираясь на интуитивные эвристики" Для сложных задач: "Проанализируй проблему подробно, учитывая все факторы" #### Ожидаемые результаты от применения:

- Более эффективное использование LLM для разных типов задач
- Сокращение избыточных рассуждений в простых вопросах
- Повышение точности в сложных задачах
- Лучшее понимание ограничений и возможностей модели

Хотя дообучение дает более значительные результаты, базовые принципы исследования вполне применимы в стандартном чате и могут существенно улучшить взаимодействие с LLM.

**## Анализ практической применимости: Разделение режимов мышления** - Прямая применимость: Высокая. Пользователи могут формулировать запросы к LLM, указывая предпочтительный тип мышления: "Ответь быстро и интуитивно" или "Проанализируй шаг за шагом". - Концептуальная ценность: Высокая. Понимание различий между быстрым и медленным мышлением помогает пользователям выбирать оптимальный подход к разным задачам. - Потенциал для адаптации: Высокий. Концепция легко адаптируется для ежедневного использования без специальных технических знаний.

**Создание специального датасета** - Прямая применимость: Средняя. Обычные пользователи не создают датасеты, но понимание когнитивных эвристик полезно для формулирования запросов. - Концептуальная ценность: Высокая. Знание о когнитивных эвристиках (anchoring bias, halo effect и т.д.) помогает понимать, как LLM могут "мыслить" по-разному. - Потенциал для адаптации: Средний. Принципы создания датасета могут быть адаптированы для создания персонализированных примеров обучения LLM.

**Выравнивание (alignment) моделей** - Прямая применимость: Низкая. Обычные пользователи не могут самостоятельно обучать модели. - Концептуальная ценность: Средняя. Понимание процесса выравнивания помогает осознать, как формируется "личность" LLM. - Потенциал для адаптации: Средний. Знание о методах выравнивания может помочь в более эффективном использовании существующих моделей.

**Анализ эффективности различных режимов мышления** - Прямая применимость: Высокая. Пользователи могут выбирать режим мышления в зависимости от задачи: Система 2 для математических задач, Система 1 для задач здравого смысла. - Концептуальная ценность: Высокая. Понимание сильных и слабых сторон каждого режима мышления помогает оптимизировать взаимодействие с LLM. - Потенциал для адаптации: Высокий. Эти знания применимы ко всем существующим LLM.

**Изучение неопределенности в ответах** - Прямая применимость: Средняя. Пользователи могут запрашивать более уверенные или более обдуманные ответы в зависимости от задачи. - Концептуальная ценность: Высокая. Понимание связи между типом мышления и уверенностью в ответах помогает интерпретировать результаты. - Потенциал для адаптации: Высокий. Можно запрашивать у LLM разные уровни уверенности для разных задач.

**Prompt:**

Применение знаний о Системе 1 и Системе 2 в промптах для GPT Исследование о спектре мышления от интуитивного (Система 1) до аналитического (Система 2) предоставляет ценные инсайты для оптимизации промптов. Вот как можно применить эти знания на практике:

## Пример промпта для решения математической задачи (Система 2)

[=====] Я хочу, чтобы ты использовал аналитический, пошаговый подход (Система 2) для решения следующей математической задачи:

"В магазине продаются наборы карандашей по 12 штук и ручек по 8 штук. Школа заказала 96 пишущих предметов. Если школа заказала 3 набора карандашей, сколько наборов ручек было заказано?"

Пожалуйста: 1. Сформулируй, что дано и что требуется найти 2. Запиши все необходимые уравнения 3. Выполни вычисления шаг за шагом 4. Проверь результат 5. Запиши окончательный ответ

Я ценю подробное объяснение каждого шага твоего мышления. [=====]

## Почему это работает

Этот промпт эффективно задействует преимущества Системы 2 мышления, обнаруженные в исследовании:

**Структурированный подход:** Исследование показало, что модели, настроенные на Систему 2, превосходят в арифметических задачах (улучшение до 7.66%)

**Пошаговая декомпозиция:** Явная просьба разбить решение на этапы активирует аналитическое мышление **Призыв к проверке:** Включение этапа проверки снижает вероятность ошибок, что соответствует более тщательному подходу Системы 2

**Запрос объяснений:** Просьба объяснить каждый шаг мышления соответствует более подробному и аналитическому стилю Системы 2 ## Альтернативный промпт для задач здравого смысла (Система 1)

Если бы задача требовала быстрого интуитивного ответа, промпт можно было бы сформулировать по-другому:

[=====] Дай быстрый, интуитивный ответ (Система 1) на следующий вопрос:

"Что произойдет, если оставить мороженое на солнце?"

Отвечай кратко, прямо и уверенно, без лишних объяснений. Мне нужен твой первый, наиболее очевидный ответ. [=====]

## Ключевые принципы использования исследования

- Выбор подхода по типу задачи: Аналитический (Система 2) для математики и логики; интуитивный (Система 1) для здравого смысла

- Явное указание желаемого стиля мышления в промпте
- Управление уровнем детализации: Больше шагов для Системы 2, краткость для Системы 1
- Адаптация к сложности задачи: Использование промежуточных подходов для задач средней сложности

Такой подход позволяет "настраивать" стиль мышления модели без дополнительной тренировки, просто через формулировку промпта.