

# ChronoSense: Исследование временного понимания в больших языковых моделях с интервалами времени событий

Дата: 2025-01-06 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2501.03040>

Рейтинг: 72

Адаптивность: 85

## Ключевые выводы:

Исследование ChronoSense направлено на оценку способности больших языковых моделей (LLM) понимать временные отношения между событиями. Результаты показывают, что современные LLM испытывают значительные трудности с временным мышлением, особенно с определением сложных временных отношений по Аллену и выполнением временной арифметики. Модели также демонстрируют признаки опоры на запоминание, а не на чистое рассуждение.

## Объяснение метода:

Исследование предоставляет готовые шаблоны запросов о временных отношениях, демонстрирует эффективность Chain-of-Thought для временной арифметики и выявляет ограничения моделей. Концепции временных отношений Аллена и стратегии промптинга применимы для повседневных запросов о хронологии, планировании и анализе исторических данных.

## Ключевые аспекты исследования: 1. **ChronoSense** - новый бенчмарк для оценки понимания временных отношений в LLM, фокусирующийся на 13 отношениях Аллена (before, after, during и т.д.) между временными интервалами событий.

**Временная арифметика** - бенчмарк включает задачи, требующие арифметических вычислений с датами: определение конечной точки события, следующего появления повторяющегося события и проверка активности события в промежуточное время.

**Абстрактные vs реальные события** - исследование сравнивает способность моделей работать с абстрактными событиями ("Событие А") и реальными историческими событиями из WikiData, выявляя влияние запоминания.

**Различные стратегии промптинга** - эксперименты с zero-shot, few-shot и chain-of-thought (CoT) подходами, демонстрирующие значительные улучшения при использовании CoT для задач временной арифметики.

**Сравнительный анализ моделей** - тестирование семи современных LLM показывает различия в их способности обрабатывать разные типы временных отношений.

## Дополнение:

### Применение методов в стандартном чате

Для работы методов этого исследования **не требуется** дообучение или API. Все подходы можно применить в стандартном чате с LLM. Ученые использовали API только для систематического тестирования разных моделей.

### Концепции и подходы для стандартного чата:

**Использование структурированных запросов о временных отношениях**

Применение: Любой пользователь может использовать предложенные в исследовании 13 шаблонов запросов (таблица 1) для формулировки вопросов о временных отношениях между событиями. Результат: Более точные ответы о хронологических связях между событиями

**Chain-of-Thought промптинг для временных вычислений**

Применение: Попросить модель "рассуждать шаг за шагом" при расчете дат, длительности или повторяющихся событий. Результат: Значительное повышение точности (с ~30-60% до 80-90% по данным исследования)

**Формулировка запросов с явным указанием временных интервалов**

Применение: Четко указывать начальные и конечные даты событий в запросах. Результат: Более точные ответы о временных отношениях

**Избегание сложных временных отношений**

Применение: Переформулировать запросы, избегая отношений типа "equals", "finishes" и "overlapped by", которые модели обрабатывают хуже. Результат: Снижение вероятности ошибочных ответов

**Учет различий в обработке симметричных отношений**

Применение: Формулировать запросы, используя отношения "before" вместо "after", "contains" вместо "during". Результат: Повышение точности ответов. ## Анализ практической применимости: 1. **Бенчмарк временных отношений** - Прямая применимость: Средняя. Пользователи могут использовать шаблоны вопросов для формулировки запросов о временных отношениях между событиями. - Концептуальная ценность: Высокая. Понимание 13 типов отношений Аллена даёт пользователям структуру для формулировки точных запросов о временных связях. -

Потенциал для адаптации: Высокий. Пользователи могут адаптировать шаблоны для запросов о хронологии событий, планировании и анализе исторических данных.

**Временная арифметика** Прямая применимость: Высокая. Пользователи могут применять продемонстрированные подходы для расчёта дат, длительности и повторяющихся событий. Концептуальная ценность: Средняя. Понимание способности и ограничений LLM в выполнении временных вычислений помогает формулировать запросы. Потенциал для адаптации: Высокий. Методы применимы для планирования, управления календарём, расчёта сроков проектов.

### **Chain-of-thought промптинг**

Прямая применимость: Очень высокая. Исследование наглядно демонстрирует эффективность CoT для временных задач, что можно сразу применять. Концептуальная ценность: Высокая. Показывает, как пошаговое рассуждение улучшает точность временных вычислений. Потенциал для адаптации: Высокий. Подход может быть адаптирован для любых временных запросов.

### **Абстрактные vs реальные события**

Прямая применимость: Низкая. Это больше исследовательский аспект. Концептуальная ценность: Высокая. Пользователи осознают влияние предварительного обучения модели на точность ответов о событиях. Потенциал для адаптации: Средний. Понимание этого аспекта помогает формулировать запросы с учётом возможной предвзятости модели.

### **Сравнение моделей**

Прямая применимость: Средняя. Помогает выбрать подходящую модель для временных задач. Концептуальная ценность: Средняя. Даёт представление о различиях в способностях моделей. Потенциал для адаптации: Низкий. Эта информация быстро устаревает с появлением новых моделей.

## **Prompt:**

Использование знаний из исследования ChronoSense в промптах для GPT ##  
Ключевые выводы для составления промптов

Исследование ChronoSense показывает, что большие языковые модели имеют определенные ограничения в понимании временных отношений. Эти знания можно использовать для оптимизации промптов при работе с временными данными.

## Пример эффективного промпта для временной задачи

[=====] # Промпт для решения задачи с временными интервалами

Я хочу, чтобы ты помог мне определить последовательность событий и их временные отношения для планирования проекта.

## Контекст Мне нужно определить, когда задача Б должна быть запланирована относительно задачи А.

## Инструкции 1. Задача А начинается 15 июня в 9:00 и заканчивается 18 июня в 17:00 2. Задача Б требует 2 полных рабочих дня (с 9:00 до 17:00) 3. Задача Б должна начаться после завершения задачи А

## Формат решения Используй пошаговое рассуждение (chain-of-thought): - Сначала определи, когда точно заканчивается задача А - Затем рассчитай, когда может начаться задача Б - Далее определи продолжительность задачи Б - Наконец, укажи конкретные даты и время начала и окончания задачи Б

Представь результат в виде календарного плана с указанием точных временных интервалов для обеих задач. [=====]

## Почему этот промпт эффективен

Данный промпт использует несколько ключевых выводов из исследования ChronoSense:

**Использует простые временные отношения** ("после") вместо сложных отношений Аллена, так как исследование показало, что модели лучше понимают базовые отношения "before" и "after".

**Применяет chain-of-thought подход**, который, согласно исследованию, значительно улучшает производительность в задачах временной арифметики (с 0.45 до 0.92 для расчета конечной точки времени).

**Предоставляет четкую структуру** для ответа, что помогает модели следовать логическому процессу рассуждения.

**Использует конкретные временные точки** вместо абстрактных событий, что снижает когнитивную нагрузку на модель.

Исследование подтверждает, что такой структурированный подход с пошаговым рассуждением значительно повышает точность ответов GPT в задачах, связанных с временными расчетами и планированием последовательностей событий.