

Chain of Draft (CoD): Минималистичные рассуждения для эффективности

Что такое Chain of Draft?

Chain of Draft (CoD) — это инновационная техника промпт-инжиниринга, позволяющая языковым моделям генерировать **минималистичные, но информативные промежуточные рассуждения**. Ключевая особенность метода в том, что он достигает точности, сравнимой с классическим Chain of Thought (CoT), но использует всего около 8% токенов, что значительно снижает стоимость и время генерации ответов.

Ключевые принципы для применения в промтах

1. **Ограничение длины промежуточных рассуждений** — указание модели использовать не более 5 слов для каждого шага мышления
2. **Сохранение пошагового подхода** — несмотря на краткость, сохраняется полная последовательность шагов рассуждения
3. **Финальное заключение в полном формате** — после кратких промежуточных шагов следует полноценный окончательный ответ
4. **Эксплицитная инструкция** — четкое указание модели на формат и объем каждого шага
5. **Итоговая сверка** — использование финального шага для проверки результатов минималистичных рассуждений

Практические примеры применения

Пример 1: Решение математической задачи

[=====]

Реши следующую математическую задачу, используя подход Chain of Draft. Для каждого шага рассуждения используй не более 5 слов.

Задача: В магазине продаются футболки по 800 рублей и джинсы по 2400 рублей. Маша купила 3 футболки и 2 пары джинсов. Сколько всего денег она потратила?

Инструкция: Думай пошагово, но записывай только краткие черновые мысли для каждого шага, максимум 5 слов на шаг. После рассуждений дай финальный ответ.
[=====]

Ожидаемый ответ:

Футболки: 3×800 руб.
Стоимость футболок: 2400 руб.
Джинсы: 2×2400 руб.
Стоимость джинсов: 4800 руб.
Общая сумма: $2400 + 4800$.
Итого: 7200 рублей.

Окончательный ответ: Маша потратила 7200 рублей.

Пример 2: Анализ текста

[=====]

Проанализируй следующий отрывок из литературного произведения, используя подход Chain of Draft. Для каждого шага анализа используй не более 5 слов.

[ТЕКСТ ДЛЯ АНАЛИЗА]

Инструкция: Проведи анализ по следующим аспектам:

1. Основная тема
2. Ключевые персонажи
3. Литературные приемы
4. Эмоциональная тональность
5. Возможная интерпретация

Для каждого аспекта анализа используй не более 5 слов в промежуточных рассуждениях. После этого дай полный развернутый анализ текста.

[=====]

Пример 3: Решение логической задачи

[=====]

Реши следующую логическую задачу, используя подход Chain of Draft. Для каждого шага рассуждения используй не более 5 слов.

Задача: В комнате находятся 4 человека: Анна, Борис, Виктор и Галина. Известно, что:
- Тот, кто сидит напротив Анны, работает врачом

- Борис сидит слева от юриста
- Программист и учитель сидят напротив друг друга
- Галина не является ни врачом, ни программистом

Определи, кто какую профессию имеет и как они расположены за столом.

Инструкция: Думай пошагово, но записывай только краткие черновые мысли для каждого шага, максимум 5 слов на шаг. После рассуждений дай финальный ответ.

[=====]

Механизм работы Chain of Draft

За счет чего работает:

1. Экономия когнитивных ресурсов

- Сокращение длины рассуждений позволяет модели фокусировать вычислительные ресурсы на ключевых аспектах задачи
- Минималистичный формат заставляет модель выделять только самую важную информацию для каждого шага

2. Структурированное мышление в компактном формате

- Несмотря на краткость, полностью сохраняется структура рассуждения
- Модель по-прежнему проходит через все необходимые логические этапы, но в сжатой форме

3. Снижение риска отклонения от темы

- Ограничение в 5 слов предотвращает "блуждание" в рассуждениях
- Повышается концентрация на ключевых аспектах решения

4. Эффект экономии контекстного окна

- Драматическое сокращение использования токенов (до 92%)
- Освобождается место в контекстном окне для более сложных задач

5. Предотвращение ошибок в длинных рассуждениях

- Исследование показывает, что ошибки часто возникают на поздних этапах длинных рассуждений
- Компактные шаги снижают вероятность таких ошибок

Практические преимущества:

- **Значительная экономия расходов** — использование всего 7.6-8% токенов по сравнению с классическим Chain of Thought
- **Сокращение времени ожидания** — быстрее генерация ответов
- **Сохранение или повышение точности** — несмотря на краткость, точность решений остается на уровне полных рассуждений
- **Лучшая масштабируемость** — можно обрабатывать более сложные и объемные задачи с теми же ресурсами
- **Применимость к разным типам задач** — метод эффективен для математических, логических, аналитических и текстовых задач

Ограничения метода

Исследование также отметило некоторые ограничения Chain of Draft:

1. **Зависимость от few-shot примеров** — метод менее эффективен при zero-shot использовании
2. **Ограниченная эффективность на малых моделях** — наилучшие результаты показывает на моделях от 3B параметров и выше
3. **Не универсальное решение** — для некоторых типов задач, требующих подробного обоснования, классический CoT может быть предпочтительнее

Сравнение с другими методами

Chain of Draft отличается от других похожих методов:

- **Стандартный Chain-of-Thought (CoT):** CoD использует те же принципы пошагового рассуждения, но с жестким ограничением на длину каждого шага
- **Chain-of-Description:** Другой метод с похожим названием, но фокусирующийся на мультимодальных задачах и требующий от модели сначала описание входных данных
- **Strategic Chain-of-Thought:** Фокусируется на предварительном планировании стратегии, а CoD — на минимализме изложения

Заключение

Chain of Draft — это простой, но чрезвычайно эффективный метод для оптимизации работы с языковыми моделями. Требуя от модели генерировать минималистичные промежуточные рассуждения (до 5 слов на шаг), метод позволяет сохранить или даже улучшить точность решений при драматическом снижении расхода токенов и времени генерации.

Этот подход особенно полезен для коммерческого использования LLM, где оптимизация затрат является важным фактором, а также для случаев, когда требуется быстрая обработка сложных задач в рамках ограниченного контекстного окна.