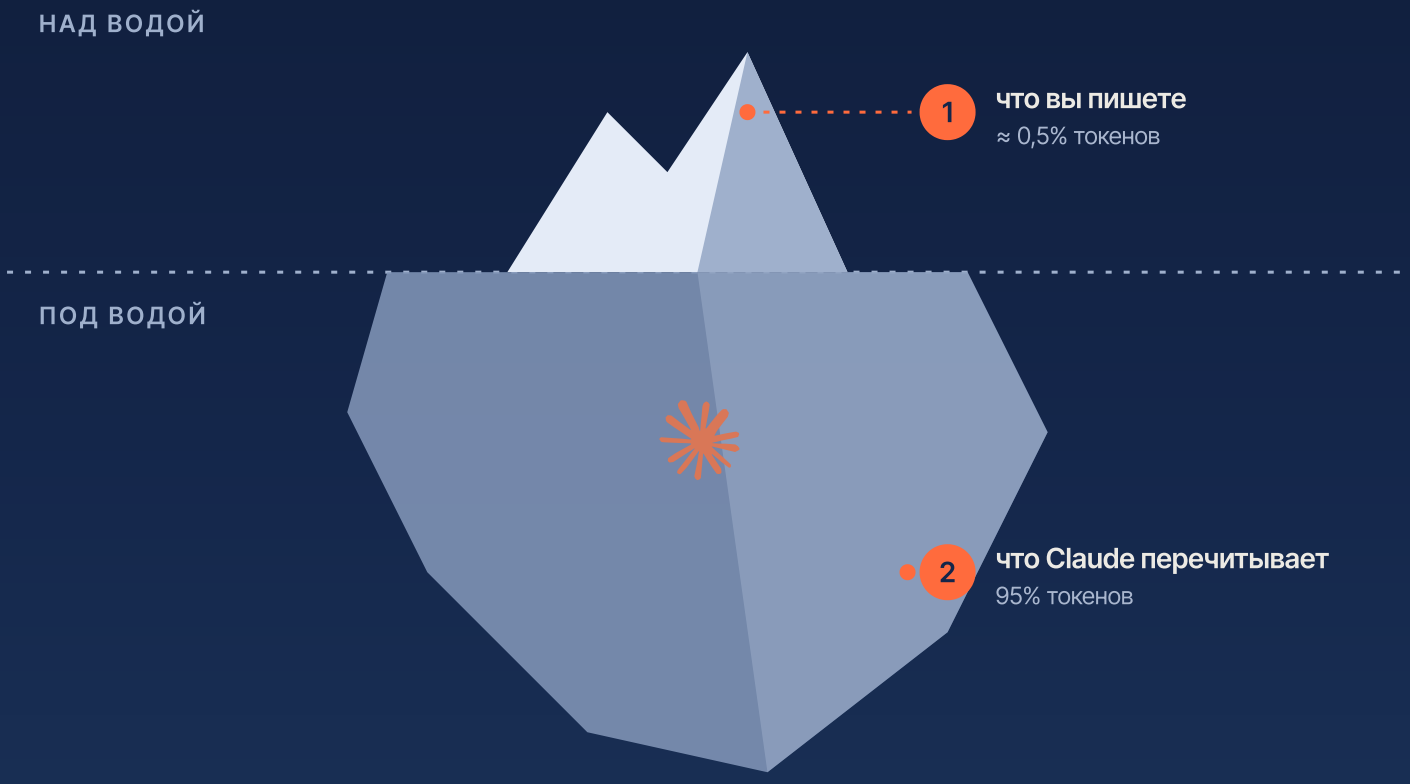
куда уходят токены claude

Вы почти ничего не пишете, а лимит кончается? Разберёмся, на что Claude тратит токены, и настроим работу так, чтобы тратилось в разы меньше.



7 правил работы с тредами · **≈ 15 минут** на настройку · **0 ₽**

главная мысль за одну минуту

У Claude нет памяти. Перед каждым вашим сообщением он заново читает всю переписку с первой страницы и только потом отвечает.

тред из 10 сообщений: Claude прочитал при новой реплике

10

тред из 300 сообщений: тот же вопрос, но читать нужно всё

300 сообщений заново, и ещё раз на следующей реплике

Как книга, которую перечитывают с первой страницы перед каждой новой строкой. Чем толще книга, тем дороже следующая строка.

ТРЕД ИЗ 10 СООБЩЕНИЙ

Читается быстро. Новая реплика почти ничего не стоит.

ТРЕД ИЗ 300 СООБЩЕНИЙ

Читается долго, и так на каждую новую реплику.

Деньги и лимиты уходят не на то, что вы написали, а на то, что Claude перечитывает.

как читать экран usage

Откройте Settings, раздел Usage. Шесть слов, и цифры станут понятны.

○	**Input** новые слова, которые вы отправили и которых Claude ещё не читал	
○	**Output** слова, которые Claude написал в ответ	
○	**Cache write** Claude записывает прочитанное в быструю память	
✓	**Cache read** достаёт прочитанное из быстрой памяти: дешевле, но тоже считается	растёт до сотен млн
○	**Threads** сколько отдельных разговоров, «комнат», было	
○	**Cache hit** доля повторного чтения из быстрой памяти	

Честная оговорка. Точный вес cache read внутри лимитов подписки Anthropic публично не раскрывает. Известно одно: чтение из быстрой памяти дешевле обычного, а объём растёт вместе с длиной треда. Поэтому мы не считаем деньги, а смотрим, где цифра растёт быстрее всего.

КАК ПРОЧИТАТЬ

Cache hit 95% значит, что почти всё, что Claude делал, это перечитывал уже прочитанное.

реальный пример: куда ушло 844 млн

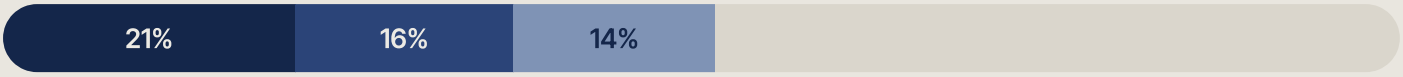
Цифры с моего экрана Usage, 2 ноября 2026. 33 треда за проект.

ВСЕГО 844,3 МЛН ТОКЕНОВ ЗА 33 ТРЕДА



оранжевая полоска: всё, что вы написали и что ответил Claude, около 4,6 млн (input 1,4 + output 3,2)

ТРИ САМЫХ ТЯЖЁЛЫХ ТРЕДА ДАЛИ 51% РАСХОДА



176,6 · 131,9 · 122,1 млн токенов, остальные 30 тредов делят оставшиеся 49%

МОДЕЛИ

Sonnet, 28 тредов	Opus, 4 треда
основная работа	135,6 млн cache read

95%
Claude перечитывал уже прочитанное

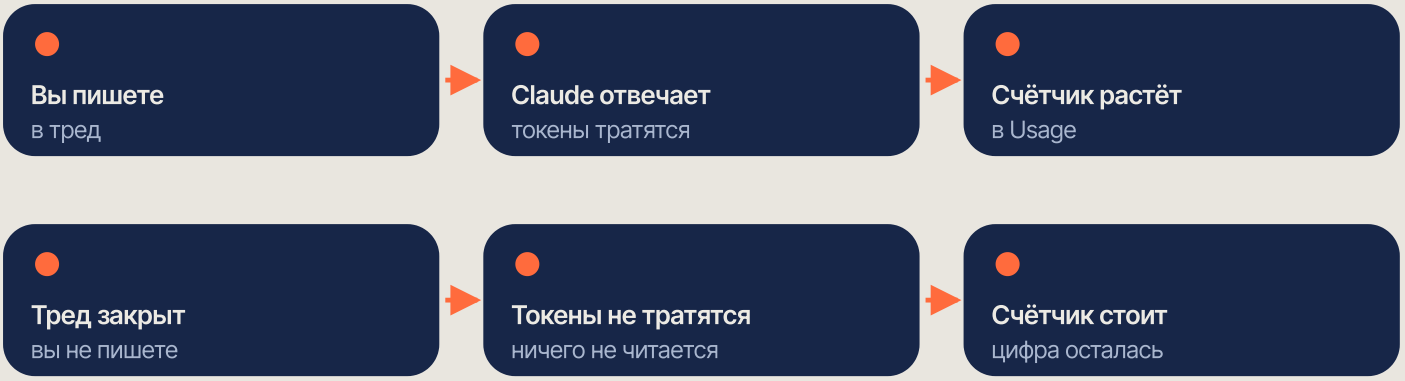
51%
расхода дали всего три треда

≈ 0,5%
это мои слова и ответы Claude

Вывод: не «я слишком много пишу», а «я слишком долго держу один тред». Три тяжёлых треда съели половину всего расхода. Ещё на экране горела красная плашка «Project instructions exceed the 16,000 character limit»: инструкции проекта были длиннее лимита. Как это исправить, расскажу на странице 8.

миф: закрытые треды тоже едят токены

Нет. Токены тратятся только в момент, когда вы пишете в тред и Claude отвечает.



Почему кажется иначе: на экране Usage итоги за всё время. Тред, который уже потратил 176 млн, останется с этой цифрой, но растёт она, только когда вы туда пишете.
Как счётчик километров в машине: он показывает, сколько проехали, а стоящая машина бензин не жжёт.

Что из этого следует

1. **Удаление само по себе не экономит**
   Старые треды удаляют для порядка, не для токенов.
2. **Экономит одно**
   Не продолжать длинные треды и не возвращаться в них.
3. **Удаляйте после проверки**
   Только когда убедились, что новый тред продолжает работу без потерь.

семь правил работы

Одна тема
один тред

→

Работа
в этом треде

→

Итог в файл
две минуты

→

Новый тред
одна строка

ПРАВИЛО 1

Одна тема, один тред

Закончили тему, закончили тред.

ПРАВИЛО 2

Итог в файл

В конце работы скажите «сохрани итог в файл». Это две минуты.

ПРАВИЛО 3

Новый тред с одной строки

«Прочитай файл такой-то, продолжаем». Не пересказывайте всё заново.

ПРАВИЛО 4

Порог 30 млн

У треда в Usage больше 30 млн токенов или тема закрыта: пора на передачу. Порог мой личный, не официальный лимит.

ПРАВИЛО 5

Сильная модель для сложного

Opus для стратегии, брендбука, больших текстов. Посты, боты и правки на Sonnet.

ПРАВИЛО 6

Одно сообщение, одна задача

Пачка мелких правок в одном сообщении дешевле, чем много отдельных.

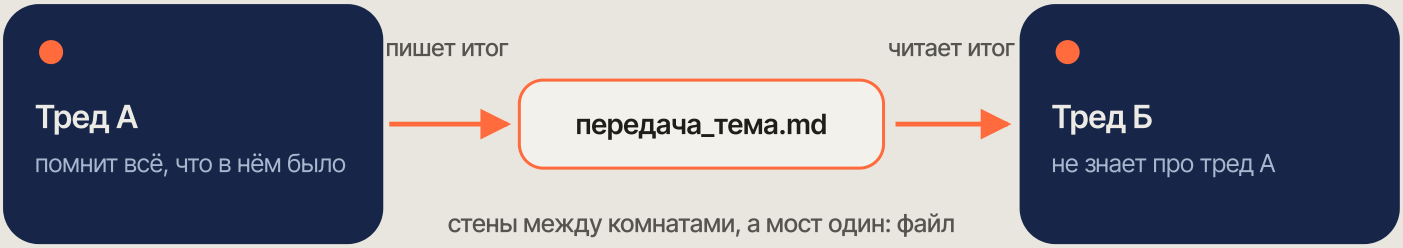
ПРАВИЛО 7

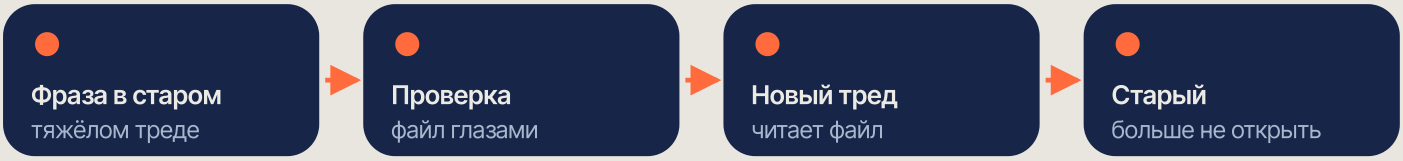
Скриншоты и большие файлы дают лишнее чтение

Можно сказать словами, говорите словами. Большие файлы читайте частями и сначала ищите нужное по ключевому слову.

передача треда: готовые фразы

Claude не видит, что сказано в другом треде. Каждый тред, это отдельная комната со своей памятью, а поиск по прошлым чатам находит только обычные чаты claude.ai, не рабочие треды в приложении. Поэтому итог должен написать сам тред: он помнит свою историю.





ШАГ 1. В СТАРЫЙ, ТЯЖЁЛЫЙ ТРЕД

Сделай файл `Задачи/Передача_[тема].md`: что решено (с датами), что в работе, ключевые файлы и ID, правила, которые мы здесь утвердили, следующие шаги. До 2 страниц, без воды.

Шаг 2. Проверьте файл глазами: всё ли важное на месте, нет ли лишнего.

ШАГ 3. В НОВЫЙ ТРЕД

Прочитай `Задачи/Передача_[тема].md`, продолжаем.

Шаг 4. Старый тред не открывайте. Через пару дней, когда новый работает, старый можно удалить или переименовать «99 на удаление».

если вылезла красная плашка

Инструкции проекта подгружаются в каждое сообщение, и лимит у них 16 000 символов. Что я сделала у себя: сумма подгружаемых файлов (общие правила, профиль, правила подпапки) была около 14 200 символов, стала около 10 400. Плюс личные настройки аккаунта, поэтому плашка ушла.



1. **Найдите, что подгружается каждый раз**
   Файлы CLAUDE.md и всё, что они подключают через @.
2. **Оставьте нужное в каждой задаче**
   Обращение, тон, главные запреты, адреса главных файлов.
3. **Редкое вынесите в отдельные файлы**
   Длинные процессы, подробный профиль, ДНК бренда, реквизиты. В основном файле одна строка-ссылка: «подробности в таком-то файле, читать, когда нужно».
4. **Сохраните копию старой версии**
   Рядом, с припиской .bak и датой.
5. **Переносите текст без изменений**
   Ничего не теряя.

Побочная польза: реквизиты и адреса больше не уезжают в каждое сообщение. Храните их в отдельном файле, который читается только по необходимости.

чек-лист на раз в три дня

Пять минут, и расход остаётся под контролем.

ДЕНЬ 1	ДЕНЬ 2	ДЕНЬ 3	ДЕНЬ 4	ДЕНЬ 5	ДЕНЬ 6	ДЕНЬ 7
чек-лист			чек-лист			чек-лист

ЧТО ПРОВЕРИТЬ

- Откройте Settings, раздел Usage
- По каждому: фраза передачи, проверка файла, новый тред
- Проверьте, не стоит ли тяжёлая модель там, где хватит лёгкой
- Найдите треды больше 30 млн токенов или старше недели
- Проверьте, что красной плашки про 16 000 символов нет

У меня раз в три дня приходит напоминание с этим чек-листом. Так не нужно помнить самой.

Что вы получаете

Тихий расход
Токены уходят на дело, а не на перечитывание.

Порядок в тредах
У каждой темы свой тред и свой файл итога.

Спокойствие
Красная плашка не застаёт врасплох.

ЧТО ДАЛЬШЕ

освободите себе вечера

Токены больше не уходят на перечитывание. Хотите
так же разгрузить остальное: ответы, сводки, документы, планы?
Тогда в голове станет тише, а вечера снова ваши.

Всё о курсе «Claude для мам за выходные» читайте на сайте,
с самого начала. Старт в декабре.

ПЕРЕЙТИ НА САЙТ

Ещё больше пользы в Telegram. Промпты, разборы
и новые гайды по нейросетям.

ОТКРЫТЬ КАНАЛ



АВТОР

Ульяна Влада. Автор системы AI Balance, которая помогла мне похудеть на 10 килограммов за 2 месяца. Без диеты и спорта. Я мама двоих детей. Настраиваю нейросети под жизнь: еда, сон, быт, работа, бизнес. Все запросы из гайда проверены мной лично, я сама использую их в повседневной жизни.

УЗНАТЬ ПРО AI BALANCE