

# Claude и длинный контекст: как работать с нейросетью в 2026 году

Разбор обновления от 13 марта 2026 — что изменилось, кому это нужно и как не переплатить

На основе реального анализа и данных бенчмарков

# Что такое Context Rot?

Context Rot — это «гниение памяти» у языковых моделей по мере роста диалога. Чем длиннее разговор, тем хуже модель помнит то, что было в середине.

- ❏ В июле 2025 года компания Chroma протестировала 18 моделей — ни одна не оказалась стабильной. Claude, ChatGPT, Gemini — все тупеют к концу диалога.

## Начало диалога

модели помнят хорошо — информация из начала сохраняется лучше всего

## Середина диалога

самая уязвимая зона — модели практически не помнят то, что было в середине

## Конец диалога

Claude держится дольше других — сдувался только ближе к лимиту

Раньше все считали это архитектурной особенностью LLM, которую невозможно исправить в ближайшие 1–2 года. **Claude изменил это 13 марта 2026 года.**

# Обновление Claude: реальные цифры

13 марта 2026 — Claude добавил 1 млн токенов в контекстное окно. Но главное не размер, а качество памяти.

**90%**

**Claude Opus — в начале контекста**

сохраняет до 90%  
изначальной информации

**78%**

**Claude Opus — через 1 млн токенов**

всё ещё помнит 78% —  
невероятный результат

**39%**

**Gemini — к середине окна**

уже к середине теряет больше  
половины контекста

## Как это измеряется

Стандартный бенчмарк — «иголки в стоге сена». Модели дают задачу найти 8 конкретных фактов внутри миллиона токенов контекста. Процент правильных ответов = качество памяти.

**GPT-4 справляется лучше Gemini, но всё равно далеко от Claude Opus и даже Claude Sonnet.**

# Claude в рейтингах Arena

Arena — независимый сервис сравнения LLM-моделей на основе реальных пользовательских оценок. Март 2026.

| Категория                             | Рейтинг                                  |
|---------------------------------------|------------------------------------------|
| Код / Vibe coding                     | все модели Anthropic занимают топ-5 мест |
| Работа с текстом                      | Claude Opus — 1-е и 2-е место            |
| Работа с документами                  | Claude — 1-е и 2-е место                 |
| Поиск                                 | Claude — 1-е место                       |
| Видео, изображения, мультимодальность | Claude не участвует — осознанный выбор   |



Anthropic осознанно сосредоточились на текстовой работе и коде. Пока ChatGPT пробует всё понемногу, Claude доминирует там, где это важно для профессионалов.

# Сравнение цен: за что платишь?

Claude стал самой дорогой моделью — и это осознанная стратегия победителя.

| Модель                       | Цена за 1 млн токенов | Примечания                         |
|------------------------------|-----------------------|------------------------------------|
| Claude Sonnet 4.6 (thinking) | \$25 / млн токенов    | Лучший выбор для вайбкодинга       |
| ChatGPT 5.2                  | \$14 / млн токенов    | Универсальный помощник             |
| Gemini 3.1 Pro               | \$12 / млн токенов    | Большой контекст, но слабая память |
| Gemini 3 Pro                 | \$12 / млн токенов    | Бюджетный вариант                  |



Даже подписка за \$200/мес может упереться в лимиты при интенсивной работе. Перед большой задачей оцени: оправдана ли стоимость?

Тот, кто выигрывает качеством — может поднимать цены. Конкурентам пока нечем догонять Claude.

# Как работать с Claude на практике

Теперь, когда понятно почему Claude лучший — разберём, как использовать это в реальной работе.

## Когда брать Claude

- Длинный профессиональный проект
- Вайбкодинг и работа с кодом
- Анализ документов и скриншотов
- Задачи, где важно не терять логику

## Какую модель выбрать

- Opus — для самых сложных задач
- Sonnet — баланс цены и качества
- Не рассчитывай на «миллион токенов»  
как на магию — важно качество памяти, а  
не размер окна

# Что изменилось на практике

Теперь можно дольше не  
переходить в новый  
диалог

Claude лучше держит  
середину и конец  
длинной задачи

Можно загружать сильно  
больше материалов в  
один поток работы

## На что всё равно нельзя надеяться

→ Модель всё ещё не идеальна

→ Даже Claude не помнит 100% контекста

→ Чем длиннее и сложнее задача, тем  
важнее всё равно проверять ответ

→ Не стоит думать, что проблема исчезла  
полностью

## Как безопасно работать

- Держи одну тему — один диалог
- Не смешивай код, аналитику, тексты и бытовые вопросы в один чат
- Если задача меняет направление, лучше явно написать модели: что уже сделано, что важно не забыть, какая текущая цель
- Периодически проси: кратко пересказать текущий статус, перечислить ключевые решения, напомнить ограничения проекта

# Когда Claude особенно силён

Длинная работа с кодом

Проекты, где нужно  
помнить много  
промежуточных решений

Анализ документов и  
сложных текстов

Работа со скриншотами и интерфейсами

Профессиональные задачи, а не просто  
«пообщаться с ИИ»

# Когда ChatGPT — лучший выбор

Если у тебя бюджет около 20 долларов

Если ты не работаешь с ИИ  
профессионально

Если тебе нужен универсальный  
помощник «понемногу для всего»

Если задача короткая и не требует  
большого контекста

# Главный минус Claude

- Он дорогой
- Для максимальной пользы нужен большой бюджет
- Даже дорогая подписка может упираться в лимиты
- Перед большой задачей лучше понимать, оправдана ли стоимость

# Практический тест перед длинной задачей

Перед длинной задачей задай себе 5 вопросов:

1. Мне точно нужен один длинный диалог, а не несколько коротких?
2. Эта задача реально требует удержания большого контекста?
3. Я готов платить за качество больше?
4. Мне важнее качество памяти, чем мультимодальность/универсальность?
5. Я смогу по ходу работы контролировать, что модель не уходит в ошибки?

## Итоговое правило

**Для  
профессиональной  
длинной работы**

Claude сейчас один из  
лучших вариантов

**Для обычного  
повседневного  
использования**

Чаще всего подходит Chat  
GPT

**Настоящее  
преимущество  
Claude сейчас**

Не просто «миллион  
токенов», а то, что он  
лучше удерживает  
полезный контекст на  
длинной дистанции