Table of Contents

# 🎯 Руководство по управлению токенами GPT

## Обзор

Администраторы могут управлять лимитом токенов для ответов GPT-4. Токены определяют максимальную длину ответа, который может сгенерировать искусственный интеллект.

## Что такое токены?

**Токены** — это единицы измерения текста для GPT. Примерно: - 1 токен ≈ 0.75 слова на английском языке - 1 токен ≈ 0.5-1 слово на русском языке - 100 токенов ≈ 75 английских слов или 50-100 русских слов

## Административные команды

### /tokens - Просмотр настроек

/tokens

Показывает текущий лимит токенов и значение по умолчанию.

**Пример ответа:**

🔧 Настройки токенов GPT:  
  
🎯 Текущий лимит: 800 токенов  
📋 По умолчанию: 500 токенов  
  
ℹ️ Информация:  
• Токены определяют максимальную длину ответа GPT  
• Больше токенов = более длинные ответы  
• Меньше токенов = более короткие ответы  
  
Команды управления:  
• /settokens КОЛИЧЕСТВО - установить новый лимит  
• /resettokens - сброс к значению по умолчанию

### /settokens - Установка лимита

/settokens КОЛИЧЕСТВО

Устанавливает новый лимит токенов для ответов GPT.

**Параметры:** - КОЛИЧЕСТВО - число от 50 до 4000

**Примеры:**

/settokens 300 # Короткие ответы  
/settokens 800 # Средние ответы  
/settokens 1500 # Длинные ответы

**Пример ответа:**

✅ Лимит токенов обновлен:  
  
🎯 Новый лимит: 800 токенов  
  
Изменения вступят в силу для новых запросов к GPT.

### /resettokens - Сброс к умолчанию

/resettokens

Сбрасывает лимит токенов к значению по умолчанию из файла .env.

**Пример ответа:**

✅ Лимит токенов сброшен к значению по умолчанию:  
  
🎯 Текущий лимит: 500 токенов  
📋 Значение по умолчанию: 500 токенов  
  
Изменения вступят в силу для новых запросов к GPT.

## Рекомендуемые значения

### 🔹 **Короткие ответы (200-400 токенов)**

* **Применение:** Быстрые, лаконичные ответы
* **Подходит для:** Простых вопросов, базовой поддержки
* **Примерная длина:** 1-2 предложения

### 🔸 **Средние ответы (500-800 токенов)**

* **Применение:** Сбалансированные ответы (по умолчанию)
* **Подходит для:** Большинства случаев психологической поддержки
* **Примерная длина:** 2-4 предложения

### 🔶 **Длинные ответы (1000-1500 токенов)**

* **Применение:** Подробн��е, развернутые ответы
* **Подходит для:** Сложных эмоциональных ситуаций
* **Примерная длина:** 4-8 предложений

### 🔺 **Очень длинные ответы (2000+ токенов)**

* **Применение:** Максимально подробные ответы
* **Подходит для:** Особых случаев, требующих детального разбора
* **Примерная длина:** 8+ предложений

## Ограничения

* **Минимум:** 50 токенов
* **Максимум:** 4000 токенов
* **По умолчанию:** 500 токенов (настраивается в .env)

## Влияние на работу бота

### ✅ **Больше токенов:**

* Более подробные и развернутые ответы
* Лучше для сложных эмоциональных ситуаций
* Больше расход API OpenAI
* Дольше генерация ответа

### ⚡ **Меньше токенов:**

* Более краткие и сжатые ответы
* Быстрее генерация ответа
* Меньше расход API OpenAI
* Подходит для простых случаев

## Мониторинг в статистике

Текущий лимит токенов отображается в команде /stats:

📊 Статистика бота:  
  
🎯 Настройки GPT:  
• Лимит токенов: 800  
  
...

## Техническая информация

### Файлы системы

* **data/tokens.txt** - хранит текущий лимит токенов
* **.env** - содержит значение по умолчанию (MAX\_TOKENS=500)
* **config.py** - функции управления токенами
* **gpt.py** - использует лимит при запросах к GPT-4

### Применение изменений

* Изменения **вступают в силу немедленно** для новых запросов
* Текущие активные запросы используют старый лимит
* Перезапуск бота **не требуется**

### Персистентность

* Настройки сохраняются в файле data/tokens.txt
* При перезапуске бота настройки загружаются автоматически
* Если файл отсутствует, используется значение из .env

## Примеры использования

### Сценарий 1: Увеличение детализации

/tokens # Проверяем текущий лимит  
/settokens 1200 # Увеличиваем для более подробных ответов

### Сценарий 2: Экономия ресурсов

/settokens 300 # Уменьшаем для экономии API

### Сценарий 3: Возврат к стандарту

/resettokens # Сбрасываем к значению по умолчанию

## Мониторинг и оптимизация

### 📊 **Отслеживание эффективности:**

1. Используйте /stats для мониторинга
2. Анализируйте качество ответов пользователей
3. Корректируйте лимит в зависимости от потребностей

### 🎯 **Оптимальная настройка:**

* Начните с **500 токенов** (стандарт)
* Увеличьте до **800-1000** если нужны более подробные ответы
* Уменьшите до **300-400** для экономии ресурсов

### ⚠️ **Важные моменты:**

* Слишком низкий лимит может обрезать важную информацию
* Слишком высок��й лимит увеличивает расходы на API
* Оптимальное значение зависит от специфики использования бота

## Устранение неполадок

### Команды не работают

* Убедитесь, что у вас есть права администратора
* Проверьте правильность синтаксиса команды

### Изменения не применяются

* Проверьте права доступа к файлу data/tokens.txt
* Убедитесь, что директория data/ существует

### Ошибки сохранения

* Проверьте свободное место на диске
* Убедитесь в корректности прав доступа к файловой системе