Аналитика и нагрузки

В стандартный шаблон MA_Router входит модуль, который отвечает за контроль нагрузки, лимитов и доступности моделей LLM.

Основные проверки

- **Суточный лимит диалогов**: `LLMDialogCounter.canStartNewDialog(limit=25)` — ограничивает количество диалогов на пользователя.
- **Текущая нагрузка**: `LLMTracer.getActiveSessionsCount(5)` — если слишком много активных сессий, диалог откладывается.
- **Активность ИИ**: `bot.getBoolAttr("ai_is_active")` — если ИИ выключен, переводит на резервный сценарий.

Логика работы

1. **Перед стартом диалога** — проверяются лимиты и нагрузка.
2. **Если лимит превышен** — пользователю отправляется уведомление, диалог переносится.
3. **Если нагрузка высокая** — диалог откладывается на несколько секунд.
4. **При старте диалога** — увеличивается счётчик, открывается сессия в LLMTracer.
5. **При завершении** — сессия закрывается.

Метрики и отчёты

- `/aistats` для получения аналитики по текущему состоянию системы
- Пример отчёта:

```
□□Статистика LLM Tracer
□□За последние 7 дней:
• Токены входящие: 2 533 129
• Токены исходящие: 313 036
• Всего токенов: 2 846 165
• Уникальных лидов: 3
• Уникальных сессий: 29
```

• Среднее вопросов на пользователя: 9.7

• Превышений 20 запросов в сутки: 1

За всё время:

• Токены входящие: 3 861 082

• Токены исходящие: 351 484

• Всего токенов: 4 212 566

• Уникальных лидов: 3

• Уникальных сессий: 38

• Среднее вопросов на пользователя: 12.7

• Превышений 20 запросов в сутки: 1

Обновлено: 2025-10-06 18:53:52

Таймауты

Для каждого запроса к LLM настраиваются таймауты в конфигурации. При превышении времени ожидания:

- **Фиксация проблемы** — информация об ошибке отправляется через Notifier
- **Fallback** — сценарий не блокируется, автоматически запускается Fallback-скрипт
- **Продолжение работы** — пользователь получает альтернативный ответ без прерывания диалога

Таймауты защищают систему от зависаний при проблемах на стороне LLM-провайдера.

Быстрые ответы на вопросы

- **Как узнать текущую нагрузку?** — Вызвать `LLMTracer.getActiveSessionsCount(5)`.
- **Как фиксируются ошибки?** — Через Notifier и запись в LLMTracer.

Версия #1
Павел Борисов создал 2025-11-06 11:40:07 UTC
Павел Борисов обновил 2025-11-12 14:03:47 UTC