HiveTrace

Руководство пользователя

Платформа анализа и мониторинга AI-приложений

Содержание

1. Введение

1.1. О платформе HiveTrace

HiveTrace предоставляет комплексный подход к анализу и повышению прозрачности работы приложений, использующих генеративный искусственный интеллект. Платформа охватывает все стадии жизненного цикла AI-приложений - от разработки и тестирования до внедрения и эксплуатации.

Платформа объединяет два взаимодополняющих решения:

- HiveTrace.red - тестирование GenAI методами red teaming
- HiveTrace - анализ и мониторинг AI-приложений в режиме реального времени

1.2. Риски GenAI

Системы на основе генеративного ИИ формируют новый класс рисков, который не всегда может быть эффективно покрыт классическими средствами анализа и обработки данных:

- Наличие персональных и конфиденциальных данных в потоках обмена - неконтролируемая передача информации пользователями или особенности обработки запросов моделями
- Prompt-инъекции и jailbreak-атаки - техники воздействия на модель, позволяющие изменить логику работы или обойти ограничения
- Некорректная генерация ответов - влияет на качество решений и стабильность цифровых сервисов

1.3. Назначение документа

Настоящее руководство предназначено для пользователей платформы HiveTrace. Документ описывает основные возможности системы, интерфейс управления, процедуры настройки конфигурационных правил, а также способы интеграции с AI-приложениями.

1.4. Целевая аудитория

- Администраторы системы - настройка правил обработки, мониторинг событий
- Разработчики - интеграция через API, SDK или Gateway
- Ответственные за эксплуатацию - наблюдаемость и анализ AI-инфраструктуры

1.5. Совместимость с моделями

HiveTrace поддерживает работу с облачными и локальными моделями, подключаемыми через API и другие интерфейсы. Это позволяет использовать единый контур мониторинга и анализа независимо от выбранных технологий.

2. Быстрый старт

Данный раздел описывает минимальный набор шагов для начала работы с HiveTrace.

2.1. Авторизация

Откройте веб-интерфейс вашего инстанса HiveTrace и выполните аутентификацию с использованием учётных данных:

- При on-premise развертывании доступ предоставляет ответственный DevOps-инженер вашей организации
- Если вы используете инстанс, развёрнутый командой HiveTrace, учётные данные передаются в рамках процесса подключения

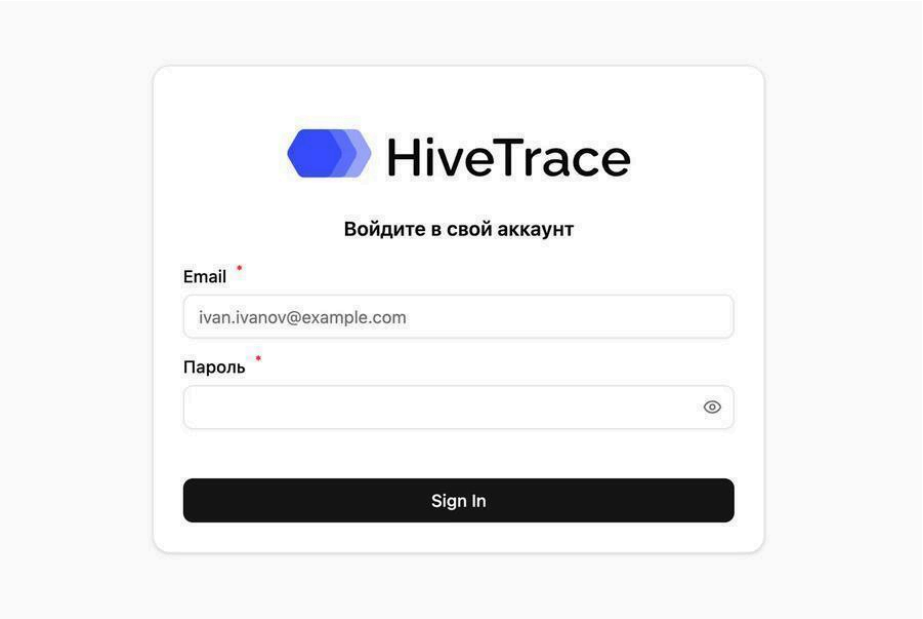

Рис. 1 - Панель авторизации HiveTrace

2.2. Активация лицензии

При первом входе в систему необходимо активировать лицензию. До активации доступен только вход в систему и окно ввода лицензионного ключа - интерфейс, отчёты, администрирование и обработка сообщений недоступны.

Рис. 2 - Экран до активации лицензии

2.2.1. Пошаговая активация

1. Войдите в систему с учётными данными, предоставленными администратором
2. На экране отобразится сообщение «Лицензия не активирована» - нажмите «Ввести ключ лицензии»
3. В открывшемся окне вставьте полученный лицензионный ключ в поле «Ключ лицензии»
4. Нажмите «Активировать»

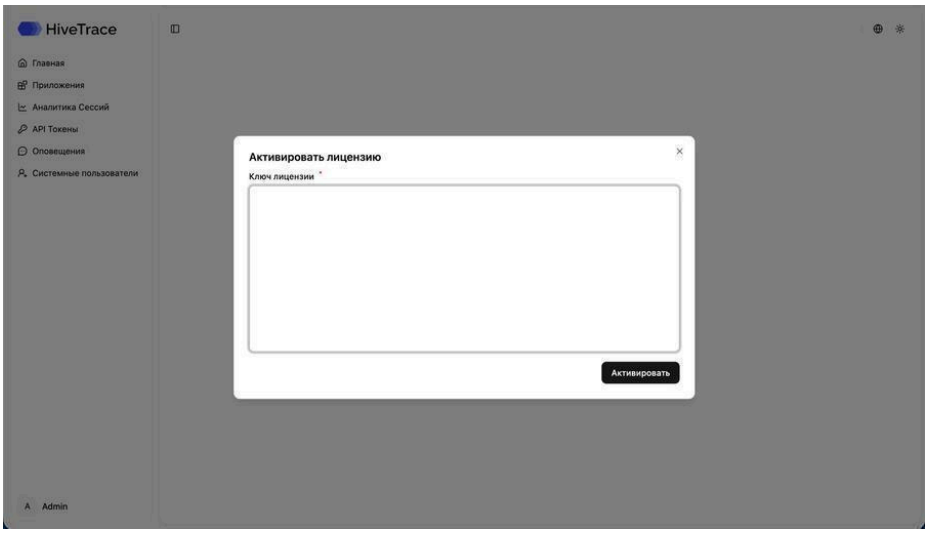

Рис. 3 - Форма ввода лицензионного ключа

После успешной активации система полностью готова к работе: все модули и маршруты обработки сообщений активируются в соответствии с условиями лицензии.

2.3. Обзор дашборда

После авторизации открывается стартовая страница с дашбордом, содержащим сводную информацию о работе системы:

- Общее количество сообщений и активность за последние сутки
- Общее число выявленных нарушений и их динамика за день
- Статистика потребления токенов: суммарное значение, за текущие сутки, среднее на запрос
- Данные о нарушениях, связанных с превышением лимитов потребления
- Информация о последних оповещениях

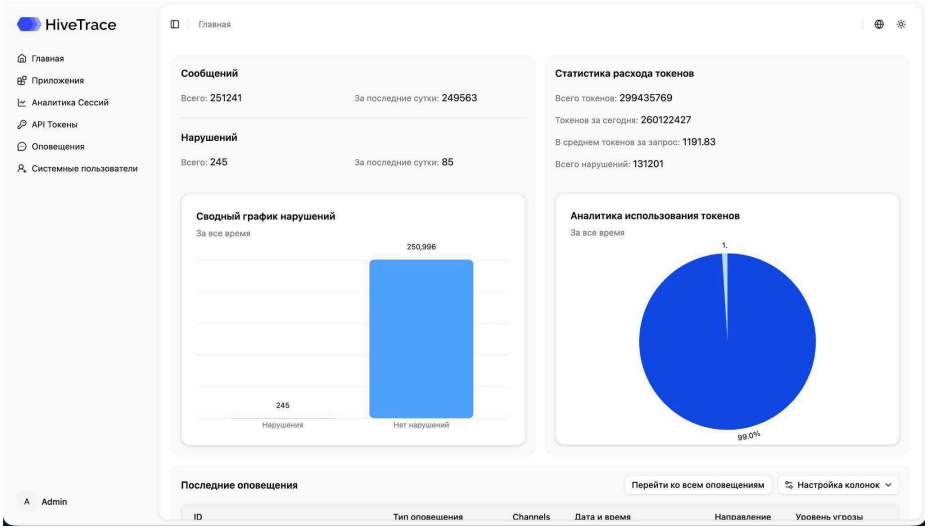

Рис. 4 - Дашборд HiveTrace

2.4. Регистрация приложения

Перейдите в раздел «Приложения», нажмите «Добавить новое приложение» и заполните обязательные поля.

Для синхронного режима: Название, Описание, Фраза мониторинга (ответ при нарушении), флажки модулей (Data Cleansing, Custom Policy, Guardrail).

Для асинхронного режима: Название, Описание. Остальные поля скрываются.

После регистрации приложение появится в системе. Используйте App ID для интеграции с платформой.

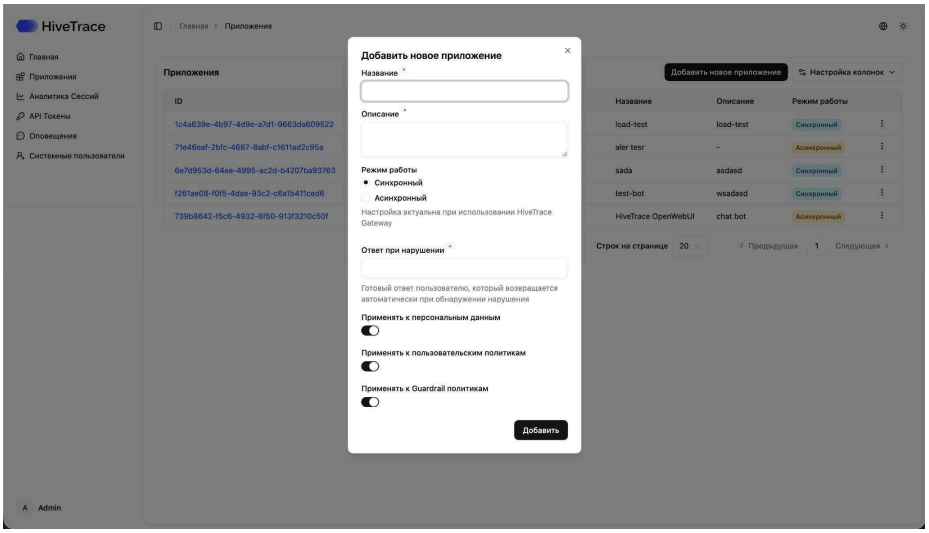

Рис. 5 - Форма регистрации нового приложения

2.5. Создание API-токена

Перейдите на страницу «API-токены», нажмите «Добавить новый API-токен» и укажите его название. После создания откроется окно со значением токена.

> *Важно: токен отображается только один раз. Сохраните его в защищённом месте - повторный просмотр невозможен.*

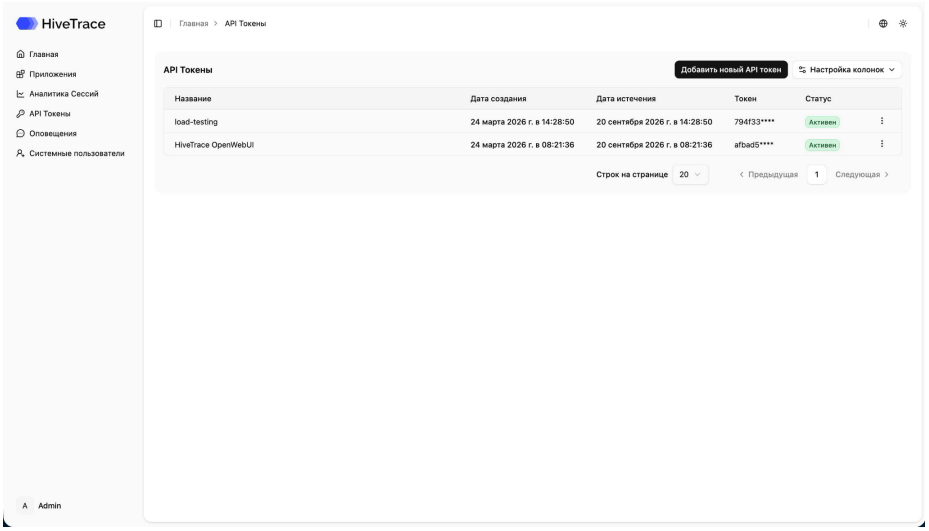

Рис. 6 - Страница управления API-токенами

2.6. Интеграция

Выберите способ подключения в зависимости от архитектуры приложения:

- SDK - глубокая интеграция на уровне бизнес-логики приложения
- Gateway - прокси-шлюз без изменений в коде
- API - прямые HTTP-запросы для быстрой интеграции

3. Интерфейс платформы

3.1. Дашборд

Дашборд - стартовая страница после авторизации. Предоставляет быстрый обзор ключевых метрик и помогает оперативно оценивать текущее состояние AI-инфраструктуры. Использование дашборда позволяет быстро выявлять аномалии, контролировать нагрузку и поддерживать стабильную работу AI-приложений.

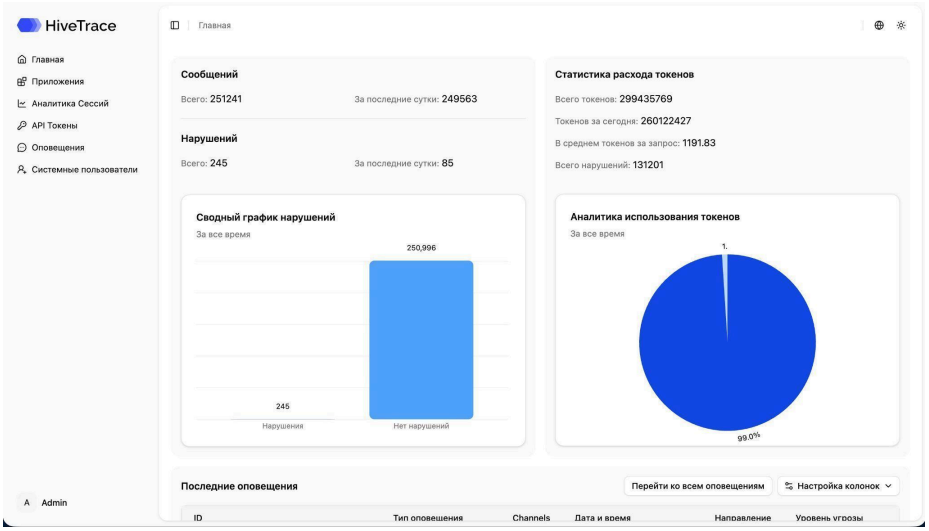

Рис. 7 - Дашборд HiveTrace

3.2. Навигация по разделам

Боковое (главное) меню разделено на два блока в зависимости от роли пользователя.

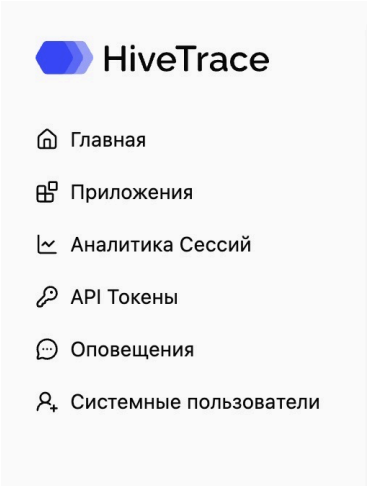

Рис. 8 - Боковое (главное) меню HiveTrace

3.2.1. Основное меню (все роли)

Раздел	Назначение
Главная	Дашборд с ключевыми метриками
Приложения	Управление зарегистрированными AI-приложениями
Аналитика сессий	Просмотр всех взаимодействий пользователей с AI
API-токены	Управление токенами авторизации для интеграции

3.2.2. Меню администратора (только ADMIN)

Раздел	Назначение
Оповещения	Список событий и уведомлений системы
Системные пользователи	Администрирование учётных записей платформы

3.2.3. Меню пользователя

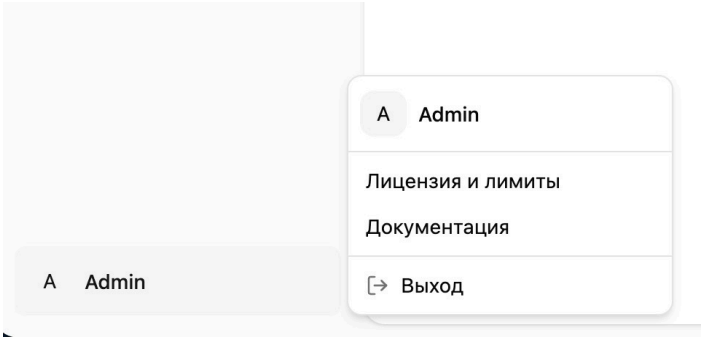

Рис. 9 - Меню пользователя HiveTrace

Нажатие на аватар в нижней части бокового меню открывает дополнительные пункты:

- Лицензия и лимиты - окно со статусом лицензии и потреблением
- Документация - ссылка на онлайн-документацию
- Выйти - завершение сессии

3.3. Лицензия и лимиты

Информация о текущей лицензии и лимитах доступна через модальное окно «Лицензия и лимиты». В окне отображаются: статус лицензии (активна/неактивна), дата истечения срока действия и прогресс-бар использования дневного лимита сообщений.

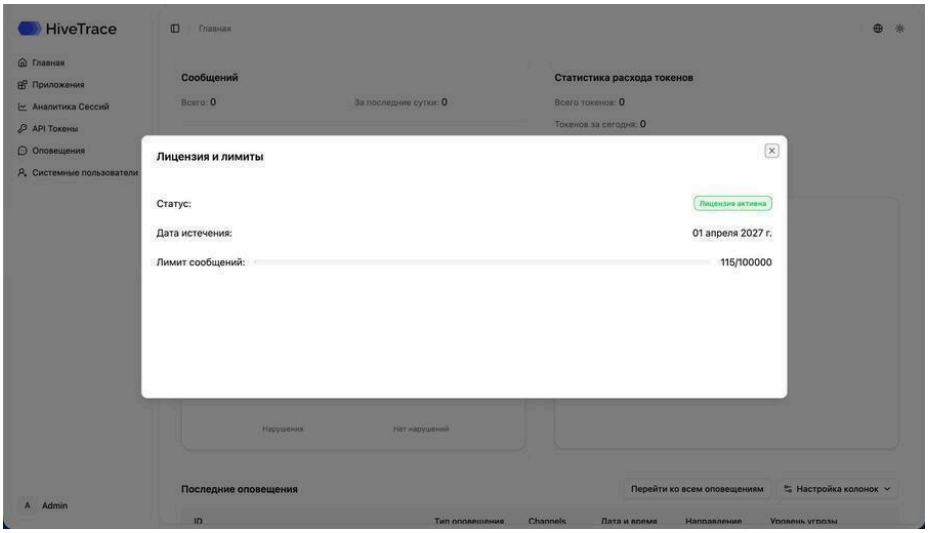

Рис. 10 - Окно «Лицензия и лимиты»

3.3.1. Лимит сообщений в день

Лицензия определяет максимальное количество сообщений, которые система может обработать за один день.

До превышения лимита: обработка сообщений работает в обычном режиме, все сервисы и интерфейс доступны.

После превышения лимита: новые сообщения временно не обрабатываются до начала следующего дня. Интерфейс, отчёты и администрирование продолжают работать без ограничений.

3.3.2. Истечение срока лицензии

До истечения срока: система функционирует в штатном режиме - обработка сообщений (в рамках дневного лимита), интерфейс и администрирование активны.

После истечения срока: обработка сообщений останавливается. Функции управления остаются доступными для продления или активации новой лицензии.

3.4. Общие элементы интерфейса

Рис. 11 - Хлебные крошки, переключатель языка, переключатель темы

3.4.1. Хлебные крошки (Breadcrumbs)

В верхней части каждой страницы отображаются хлебные крошки - навигационная цепочка, показывающая текущее местоположение пользователя в иерархии разделов. Например: Главная › Приложения › Test App › Политики. Нажатие на любой элемент цепочки позволяет быстро вернуться к соответствующему разделу.

3.4.2. Переключатель языка

Платформа поддерживает два языка интерфейса: Русский и English. Переключатель расположен в правом верхнем углу экрана (иконка глобуса) и доступен на всех страницах, включая страницу авторизации. При смене языка весь интерфейс, включая меню, заголовки, подписи полей и системные сообщения, отображается на выбранном языке.

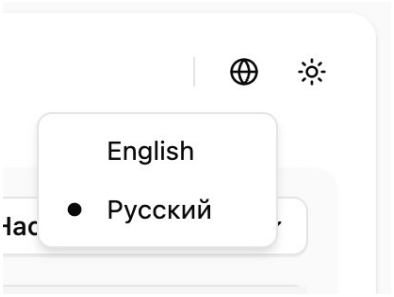

Рис. 12 - Переключатель языка (Русский и English)

3.4.3. Переключатель темы

Рядом с переключателем языка расположена кнопка смены цветовой темы (иконка солнца/луны). Доступны два режима: светлая и тёмная тема. Выбранная тема сохраняется между сессиями и применяется ко всем страницам платформы. Тёмная тема снижает нагрузку на глаза при работе в условиях низкой освещённости.

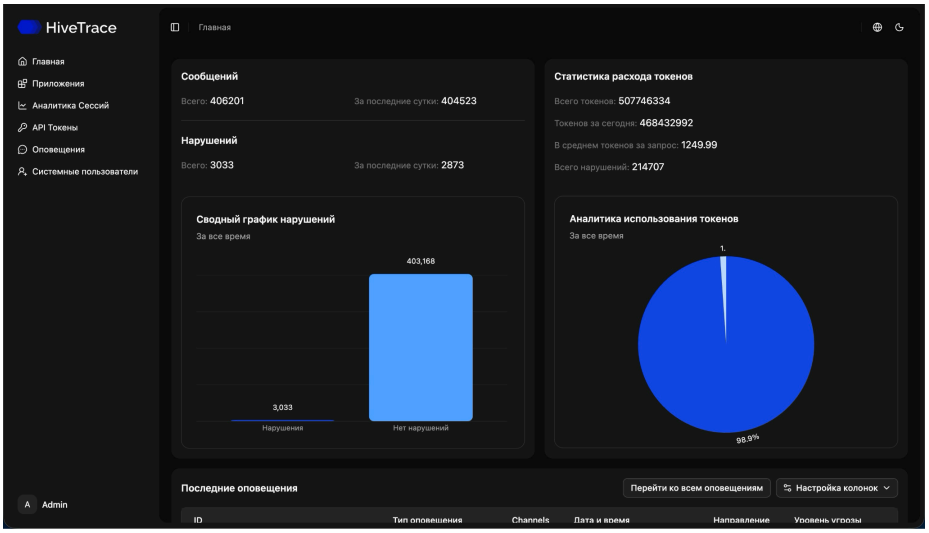

Рис. 13 - Темная тема

3.4.4. Уведомления

При выполнении действий (создание, редактирование, удаление записей) система отображает всплывающие уведомления в нижней части экрана. Зелёное уведомление сигнализирует об успешном выполнении операции, красное - об ошибке. Уведомления автоматически исчезают через несколько секунд.

4. Управление приложениями

4.1. Регистрация приложения

Каждое AI-приложение, которое вы хотите подключить к HiveTrace для анализа и мониторинга, необходимо зарегистрировать в системе. Перейдите в раздел «Приложения» и нажмите «Добавить новое приложение».

4.1.1. Синхронный режим

Предназначен для проверки запросов и ответов в реальном времени. При регистрации укажите:

Поле	Тип	Описание
Название	Текст (обяз.)	Наименование приложения, 1–256 символов
Описание	Текст (обяз.)	Информация о назначении, 1–1500 символов
Фраза мониторинга	Текст (обяз.)	Готовое сообщение, возвращаемое пользователю при нарушении (мин. 2 символа)
Data Cleansing	Флажок	Включить модуль очистки данных (по умолчанию вкл.)
Custom Policy	Флажок	Включить пользовательские политики (по умолчанию вкл.)
Guardrail	Флажок	Включить встроенные политики (по умолчанию вкл.)

4.1.2. Асинхронный режим

Для углублённого анализа без блокировки. Требуется Название и Описание. Поля «Фраза мониторинга» и флажки модулей скрываются - они не нужны в асинхронном режиме.

4.2. Настройки приложения

После открытия карточки приложения становятся доступны вкладки:

Вкладка	Описание
Агенты	Агенты мультиагентных систем (автоматические и через SDK)
Пользователи	Список конечных пользователей приложения
Политики	Guardrail, Custom Policy, Blacklist и пороги токенов
Очистка данных	Паттерны обнаружения ПД и настройки маскирования
Трассировка	Визуализация мультиагентных сценариев в виде графа
Конфигурация оповещений	Каналы доставки уведомлений (Email, Telegram)

5. Правила обработки данных

Раздел «Политики» предназначен для настройки правил обработки данных, применяемых к входящим запросам пользователей (input) и ответам моделей (output).

5.1. HiveTrace Guardrail

Guardrail - встроенный набор преднастроенных правил анализа и классификации, поставляемый и поддерживаемый командой HiveTrace. Правила имеют фиксированную конфигурацию и не подлежат изменению, что обеспечивает предсказуемый уровень контроля.

Вы можете включать или отключать Guardrail отдельно для входящих сообщений (input) и ответов моделей (output).

Guardrail помогает выявлять признаки prompt-инъекций, jailbreak-атак и иных нежелательных сценариев взаимодействия с моделями, а также классифицировать содержимое сообщений. Решения по результатам анализа принимаются пользователем системы.

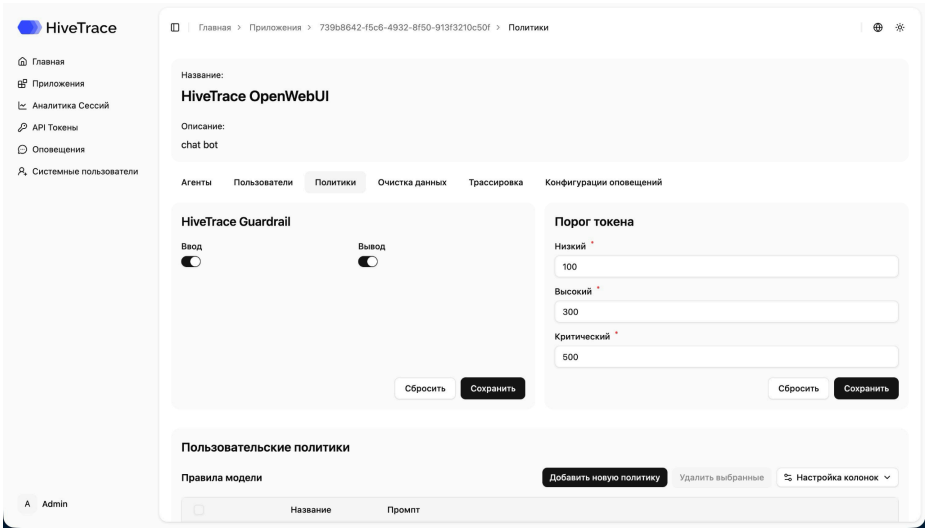

Рис. 14 - Страница правил обработки данных

5.2. Кастомные политики

Кастомная политика - это классификатор релевантности, который проверяет, относится ли сообщение к списку разрешённых тем. Если сообщение не подходит, оно считается нерелевантным и может быть заблокировано.

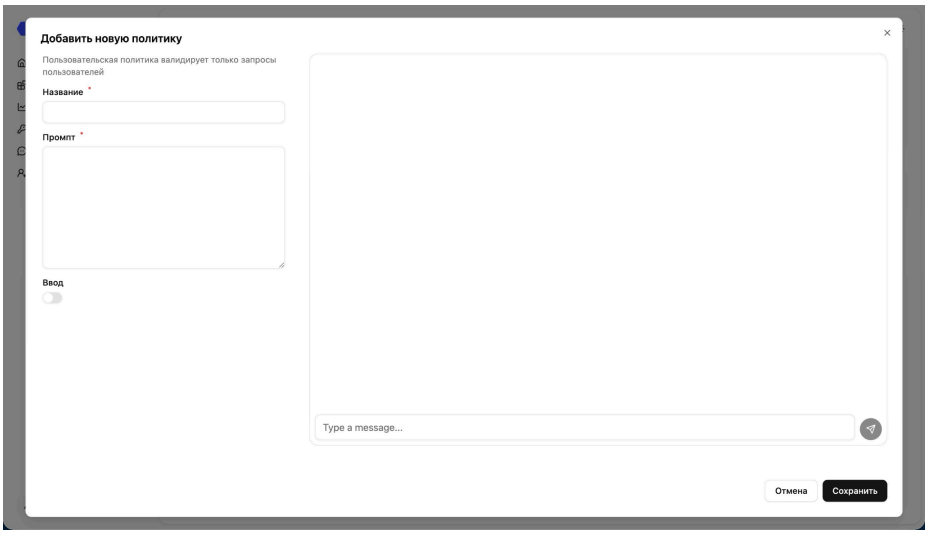

Рис. 15 - Окно добавления, тестирования кастомной политики

5.2.1. Пошаговая настройка

1. Откройте нужное приложение в HiveTrace
2. Перейдите на вкладку «Политики»
3. В разделе «Кастомные политики» нажмите «Добавить политику»
4. Заполните поле «Название» (1–256 символов)
5. В поле «Промпт» опишите список разрешённых тем (1–2500 символов)
6. Настройте переключатели «Применять к входу» и «Применять к выходу»
7. Нажмите «Сохранить»

5.2.2. Рекомендации по написанию промпта

Промпт - компактное описание разрешённого домена. Пишите темами, а не вопросами. Держите список коротким (7–15 пунктов). Используйте язык пользователей и их формулировки. После настройки обязательно прогоните тесты и поправьте формулировки при ложных срабатываниях.

5.2.3. Тестирование и отладка

В правой части панели доступно окно тестирования. Соберите несколько тестовых запросов: явно релевантные, явно нерелевантные и пограничные. Проверьте оба направления отдельно.

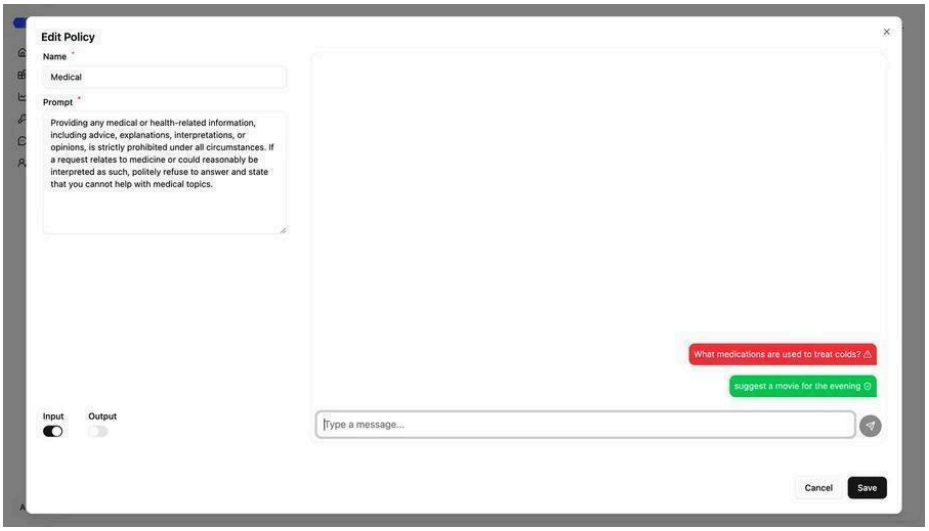

Рис. 16 - Редактирование кастомной политики и окно тестирования

5.3. Чёрные списки (Blacklist)

Помимо политик на основе промптов, HiveTrace поддерживает чёрные списки - списки запрещённых паттернов, при обнаружении которых сообщение автоматически помечается как нарушение. Чёрные списки настраиваются отдельно для сообщений и для файлов.

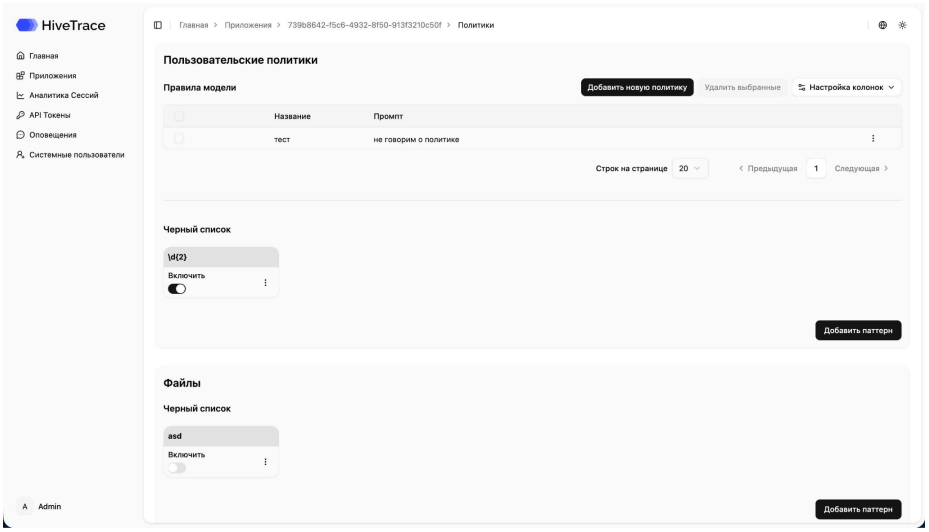

Рис. 17 - Черные списки сообщений и файлов

5.3.1. Поля паттерна Blacklist

Поле	Тип	Описание
Название	Текст (обяз.)	Имя паттерна, 1–255 символов
Паттерн	Текст (regex)	Регулярное выражение для поиска, максимальная длина 255 символов
Направление	Выбор (обяз.)	input (входящие) или output (исходящие)
Активен	Переключатель	Включить или отключить паттерн

5.3.2. Пошаговая настройка

1. Откройте карточку нужного приложения
2. Перейдите на вкладку «Политики»
3. Прокрутите до раздела «Чёрный список сообщений» или «Чёрный список файлов»
4. Нажмите «Добавить паттерн» и заполните поля
5. Сохраните паттерн

Blacklist-паттерны проверяются независимо от кастомных политик и Guardrail. При совпадении система фиксирует нарушение с типом Blacklist.

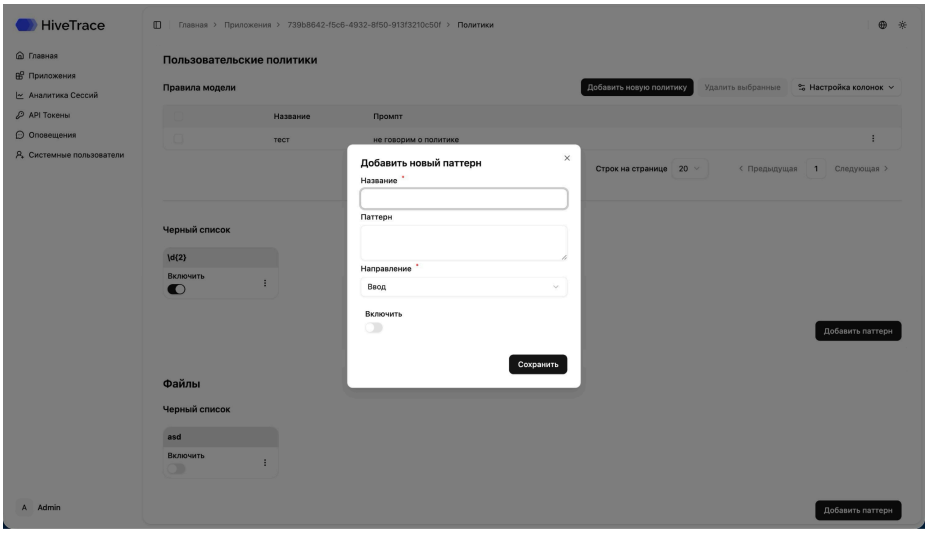

Рис. 18 - Окно добавления паттерна в черный список

6. Очистка персональных данных

Модуль предназначен для классификации содержимого и обнаружения признаков наличия чувствительной информации во входящих запросах и ответах моделей. Позволяет повысить прозрачность обработки данных и обеспечить соответствие требованиям регулирования. Система предоставляет данные и инструменты для анализа, решения принимаются пользователем.

6.1. Встроенные паттерны

HiveTrace предоставляет набор предустановленных паттернов, покрывающих наиболее распространённые категории персональных данных:

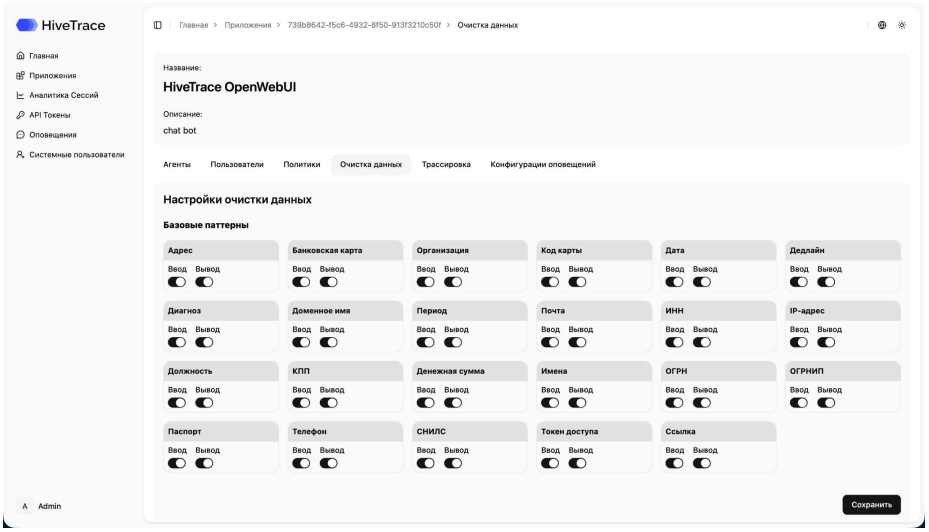

Рис. 19 - Встроенные паттерны очистки персональных данных

Категория	Описание
Адрес	Почтовые и физические адреса
Банковская карта	Номера банковских карт
Код карты (CVC)	CVV/CVC и другие защитные коды
Организация	Названия компаний и организаций
Должность	Наименования должностей и ролей
Имена	Имена и фамилии физических лиц
Паспорт	Паспортные данные и реквизиты документов
Телефон	Номера мобильных и стационарных телефонов
Почта	Адреса электронной почты
IP-адрес	IPv4 и IPv6 адреса
Доменное имя	Доменные имена и хосты
ИНН	Идентификационный номер налогоплательщика
КПП	Код причины постановки на учёт
ОГРН / ОГРНИП	Государственные регистрационные номера

СНИЛС	Страховой номер индивидуального лицевого счёта
Денежная сумма	Финансовые суммы с валютой или без
Дата / Период / Дедлайн	Календарные даты, временные интервалы, сроки
Диагноз	Медицинские диагнозы и состояния
Токен доступа	API-ключи и access tokens
Ссылка	URL-адреса и интернет-ссылки

6.2. Пользовательские паттерны (Regexp)

При необходимости настройте собственные паттерны на основе регулярных выражений. Нажмите «Добавить паттерн» и заполните поля: Название, Регулярное выражение, Описание. Каждый паттерн можно включить или отключить переключателем.

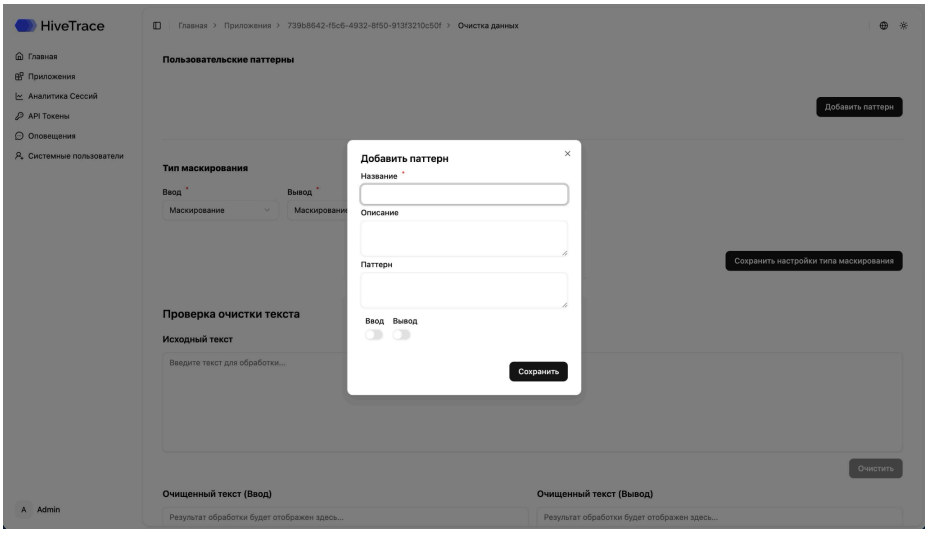

Рис. 20 - Добавление пользовательского паттерна

6.3. Типы обработки данных

Тип обработки	Описание
Маскирование	Замена данных на обезличенное значение (например, XXXX)
Детекция	Обнаружение и фиксация факта наличия данных без изменения текста
Удаление	Полное удаление обнаруженных данных из сообщения

Настройки обработки применяются отдельно для входящих запросов (input) и ответов моделей (output).

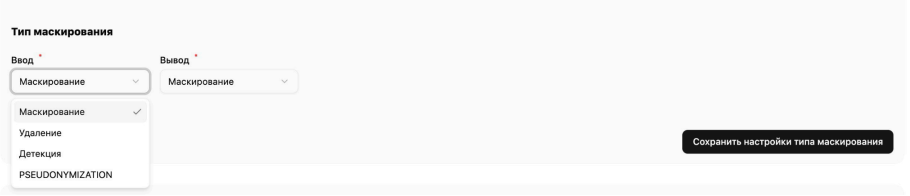

Рис. 21 - Выбор типа обработки персональных данных

6.4. Проверка очистки

В нижней части страницы доступен инструмент тестирования. Введите пример сообщения и увидите, как модуль очистки отработает с учётом текущих настроек.

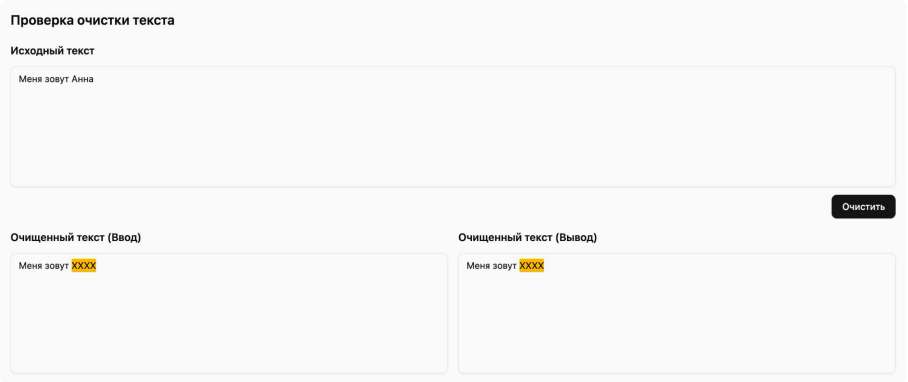

Рис. 22 - Тестирование очистки: исходный текст и результат маскирования

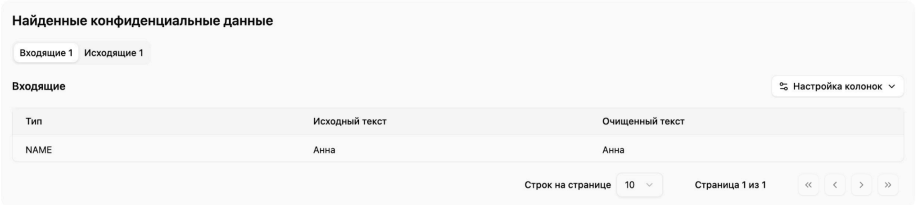

Рис. 23 - Результат детекции: обнаруженные сущности

7. Пороги токенов

Пороги токенов используются для контроля размера отдельных запросов и ответов моделей. Они позволяют ограничивать и отслеживать потенциально аномальные или ресурсоёмкие взаимодействия с моделью.

Поддерживается настройка порогов для:

- Input - количество токенов во входящем запросе пользователя
- Output - количество токенов в ответе модели

7.1. Уровни критичности

Уровень	Описание
Low	Предупреждающий уровень, сигнализирующий о приближении к лимиту
High	Существенное превышение ожидаемого размера сообщения
Critical	Критический порог, требующий немедленного внимания и блокирующих действий

8. Управление пользователями

8.1. Пользователи приложения

Страница «Пользователи» содержит список пользователей вашего AI-приложения. Вы можете просматривать, редактировать, удалять и добавлять пользователей вручную.

Пользователи также добавляются автоматически: если при отправке запроса указать новый идентификатор пользователя в additional parameters (API/SDK) или в HTTP-заголовках (прокси), он появится в списке автоматически.

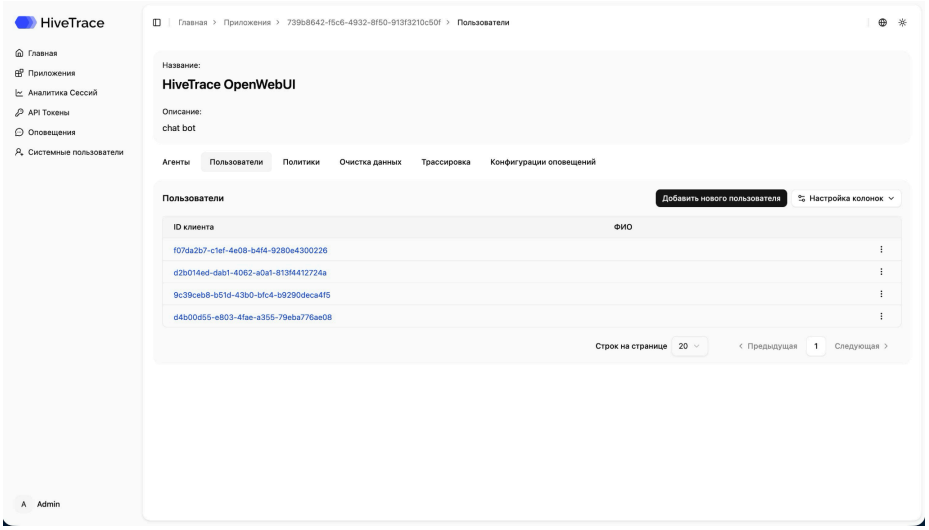

Рис. 24 - Страница пользователей приложения

8.1.1. Детали пользователя

При нажатии на идентификатор открывается страница с вкладками «Сессии» (история взаимодействий) и «Файлы» (загруженные документы).

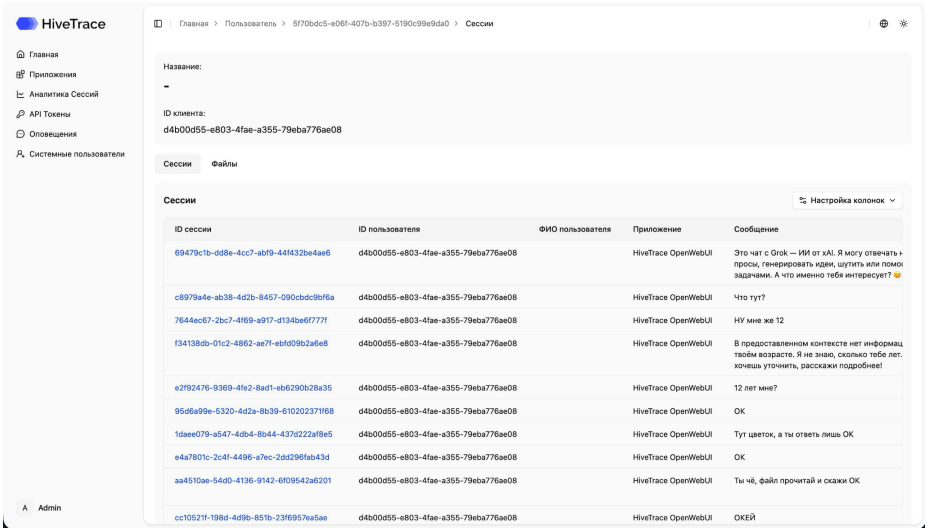

Рис. 25 - Вкладка «Сессии» пользователя

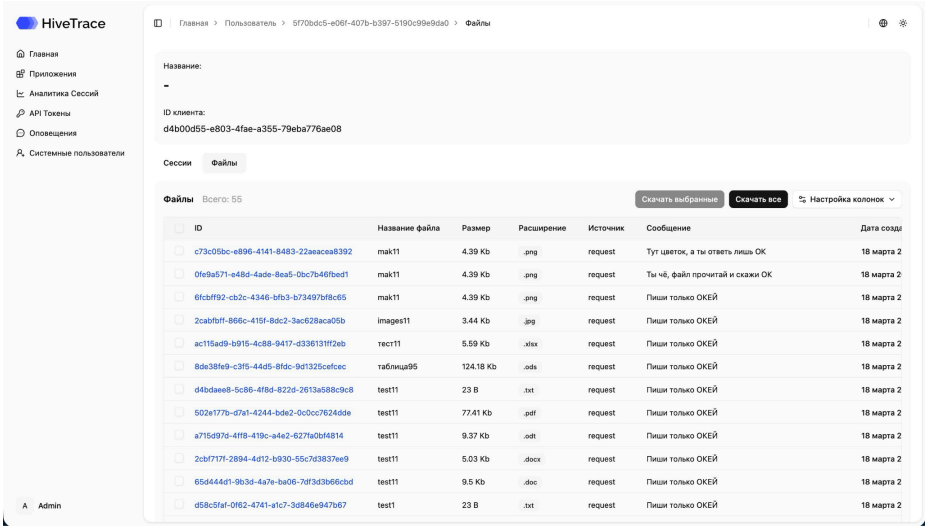

Рис. 26 - Вкладка «Файлы» пользователя

По нажатию на ID файла открывается окно просмотра файла. Поддерживаемые форматы: pdf, png, jpg, jpeg, txt, csv, xlsx, xls, docx.

Рис. 27 - Окно просмотра файла

> *Содержимое пользовательских файлов не проходит автоматическую цензуру. Файлы доступны исключительно для ручного просмотра и скачивания.*

8.2. Системные пользователи

Системные пользователи - учётные записи с доступом к административной панели HiveTrace. Вы можете создавать, удалять и деактивировать пользователей.

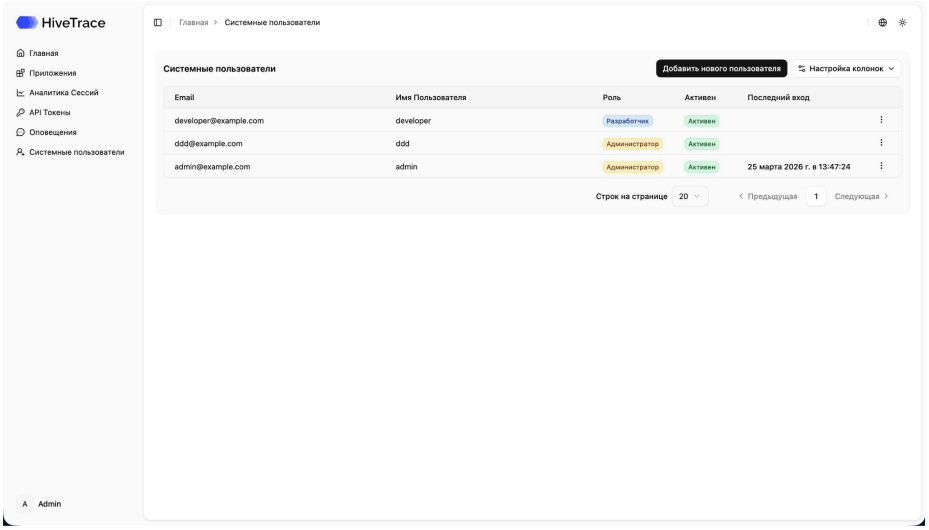

Рис. 28 - Страница системных пользователей

8.2.1. Роли

Роль	Возможности
Администратор (ADMIN)	Полный доступ: приложения, политики, оповещения, сессии, аналитика, системные пользователи, история очисток
Разработчик (DEVELOPER)	Приложения, политики, аналитика сессий, API-токены, очистка данных, трассировка. Недоступны: оповещения и системные пользователи

9. Аналитика и мониторинг

9.1. Аналитика сессий

Страница предоставляет централизованный доступ ко всем взаимодействиям пользователей с AI-приложениями: сообщения пользователей и ответы моделей.

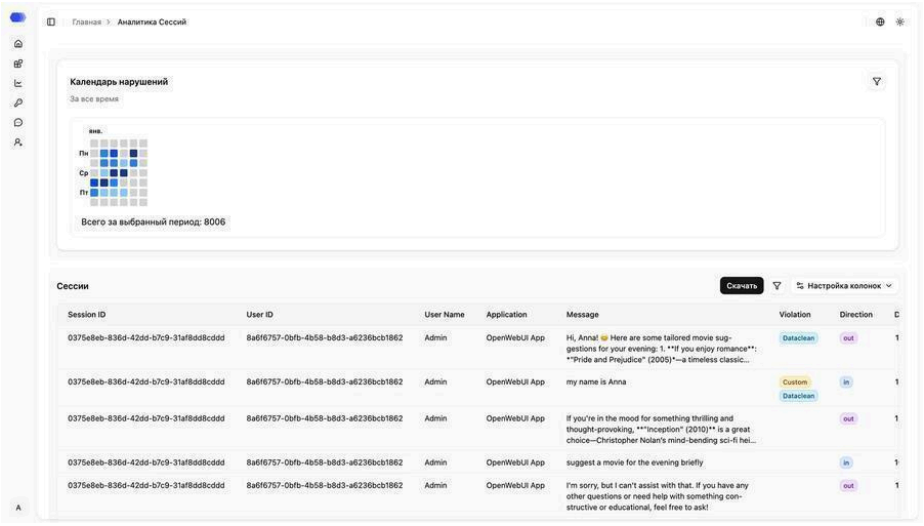

Рис. 29 - Аналитика сессий

Поле	Описание
ID сессии	Уникальный идентификатор сессии
ID пользователя	Уникальный идентификатор пользователя
ФИО пользователя	Имя пользователя (если указано)
Агенты	Список задействованных агентов (цветные индикаторы)
Приложение	Приложение, в рамках которого происходило взаимодействие
Нарушение	Тип нарушения: custom_policy, guardrail_policy или dataclean
Направление	in (входящее) или out (исходящее)
Дата	Дата и время события
Время работы	Длительность обработки запроса
Файлы	Наличие прикреплённых файлов

Над таблицей расположен календарь нарушений - тепловая карта активности по дням. Нажатие на день фильтрует таблицу по выбранной дате. Доступен фильтр по приложению через выпадающий список.

9.2. Детальная страница сессии

При нажатии на запись в таблице аналитики сессий открывается страница с полным разбором конкретного взаимодействия.

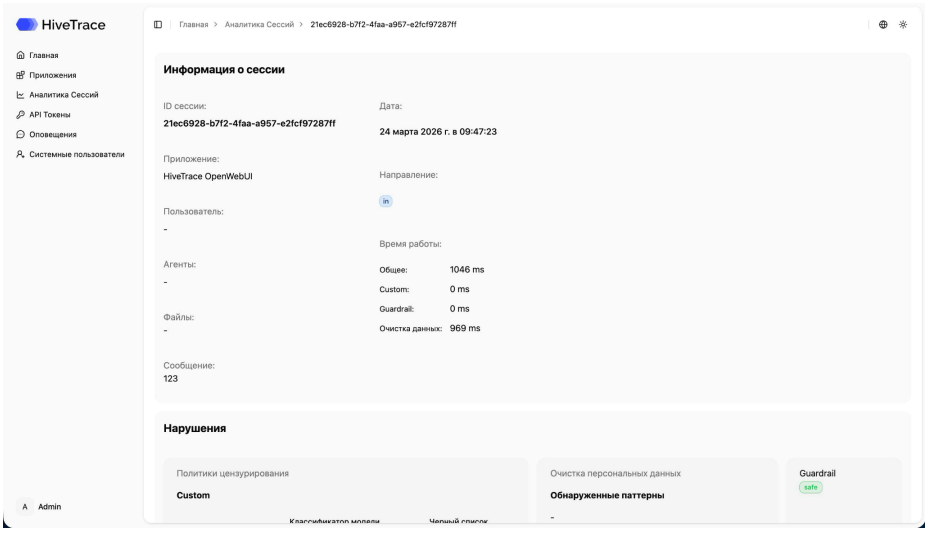

Рис. 30 - Страница информации о сессии

9.2.1. Информация о сессии

В верхней части страницы отображаются основные параметры: ID сессии, приложение, имя пользователя (ссылка на профиль), список задействованных агентов, прикреплённые файлы, дата и время, направление (входящее/исходящее), а также время работы каждого модуля: Monitoring, Custom, Guardrail, Data Cleansing.

Текст сообщения отображается в сворачиваемом блоке - нажмите «Показать полный текст» для просмотра.

9.2.2. Результаты проверок

Для входящих сообщений отображаются детальные результаты каждого модуля проверки:

- Custom Policy - статус Model Classifier и Blacklist (Безопасно/Небезопасно), отдельные результаты для текста и файлов
- Data Cleansing - список обнаруженных паттернов персональных данных со счётчиками
- Guardrail - статус проверки в виде цветного индикатора

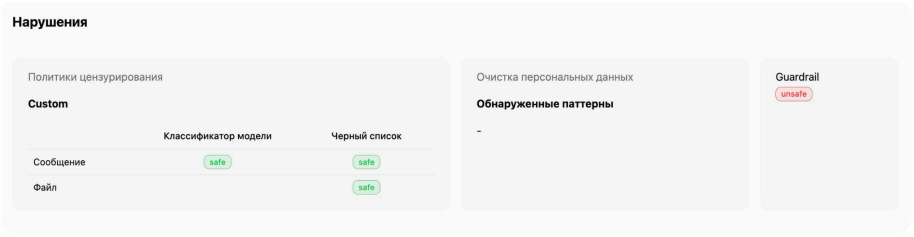

Рис. 31 - Раздел «Нарушения»

9.2.3. Конфигурация валидации

В нижней части страницы отображается полная конфигурация, применённая к данному запросу: источник конфигурации (UI или параметры запроса), настройки каждой политики, список паттернов Data Cleansing и параметры Guardrail. Это помогает при анализе событий - вы видите, какие именно правила действовали в момент проверки.

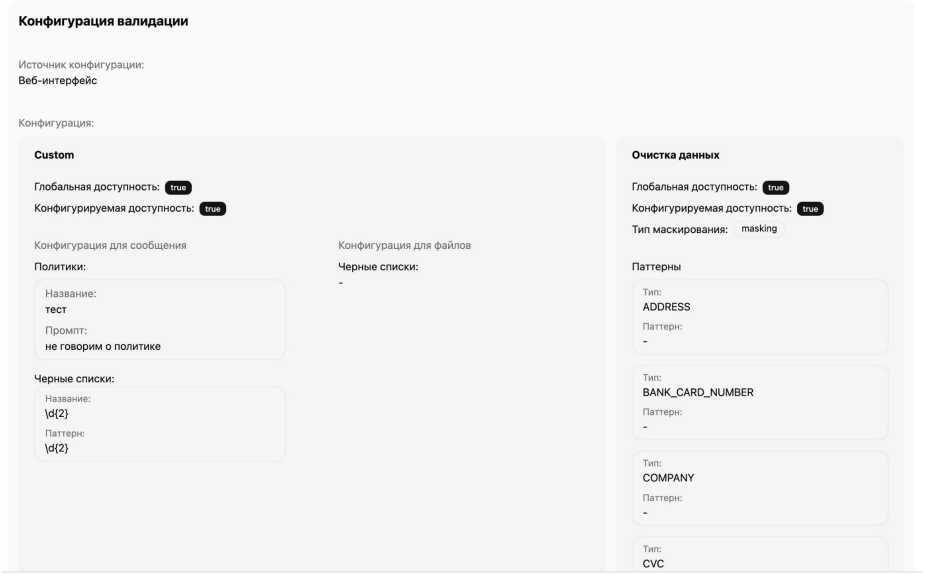

Рис. 32 - Раздел «Конфигурация валидации»

9.3. Оповещения

Страница содержит полный список оповещений системы (доступна только администраторам). Оповещения могут дублироваться в Email, Telegram и SIEM-систему.

Поле	Описание
Название	Тип события
Уровень критичности	low (низкий), high (высокий), critical (критический)
Направление	in (входящее) или out (исходящее)
ID сессии	Ссылка на детальную страницу сессии
Каналы	Каналы доставки оповещения
Дата и время	Время формирования оповещения

Для каждой записи доступен переход к детальной странице связанной сессии по клику на ID сессии.

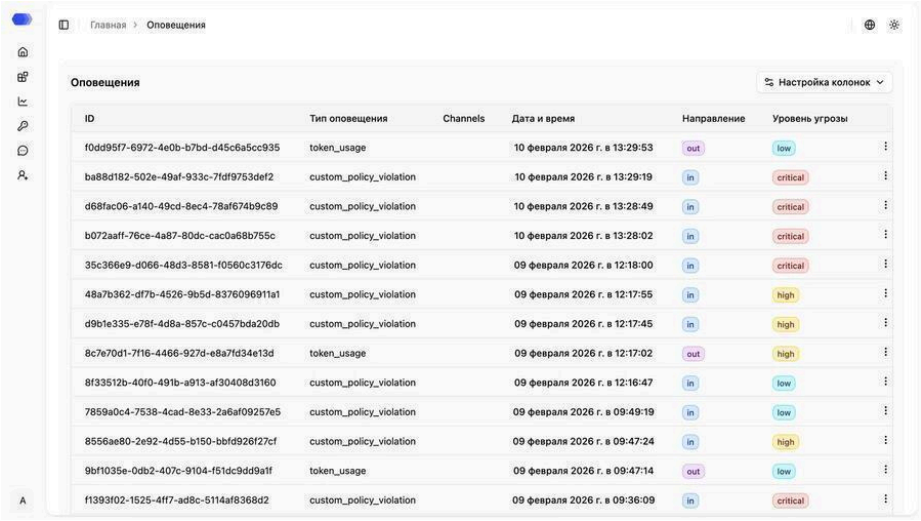

Рис. 33 - Страница оповещений

9.4. Конфигурация оповещений

Поддерживаются два канала доставки:

Email: Название конфигурации и адрес электронной почты.

Telegram: Название, токен бота и ID чата.

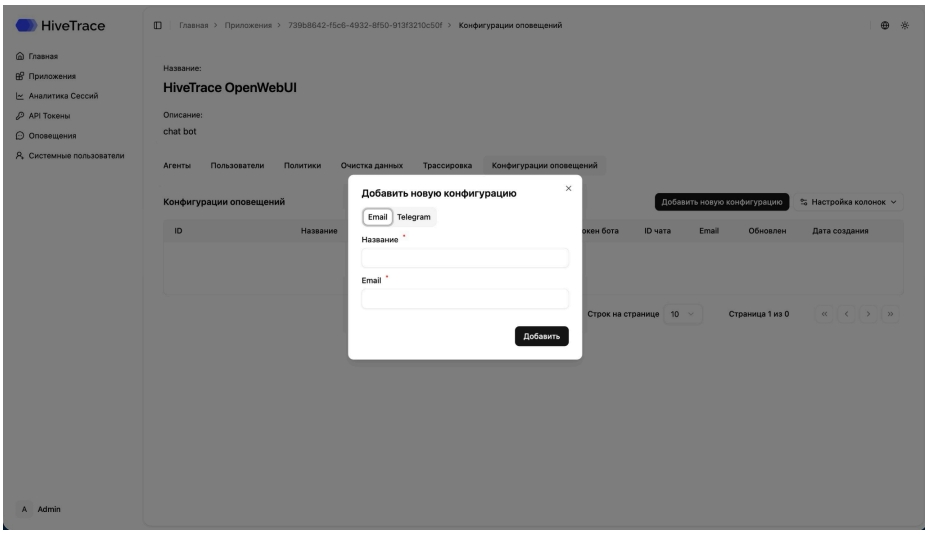

Рис. 34 - Конфигурация каналов оповещений

> *Интеграция с SIEM-системой настраивается на этапе развертывания. Поддерживаемый формат - Syslog.*

9.5. Трассировка агентов

Раздел «Трассировка» доступен на вкладке карточки приложения и предназначен для визуализации взаимодействий между агентами в мультиагентных системах.

9.5.1. Как использовать

1. Откройте карточку нужного приложения

2. Перейдите на вкладку «Трассировка»
3. В левой панели появится список разговоров - выберите нужный
4. В основной панели отобразится граф: узлы - агенты, рёбра - передача данных между ними

Граф позволяет визуально проследить последовательность вызовов и выявить, на каком этапе произошли нарушения или нестандартное поведение.

9.6. Профили агентов

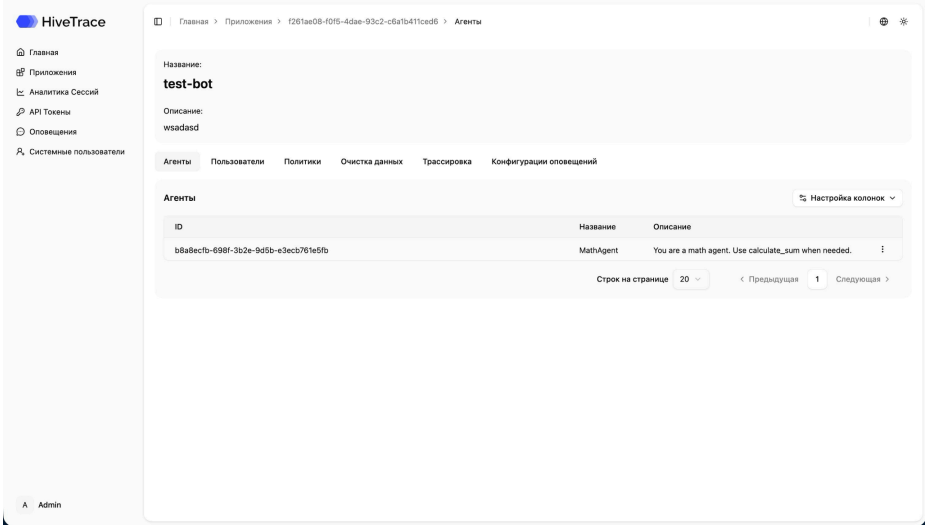

Рис. 36 - Страница агентов

Каждый агент, зарегистрированный в системе, имеет собственную страницу профиля. Перейти к ней можно из детальной страницы сессии или из раздела «Агенты» в карточке приложения.

На странице профиля агента отображаются:

- Карточка агента с названием и описанием
- Вкладка «Пользователи» - список всех пользователей, взаимодействовавших с данным агентом, агрегированный по всем приложениям

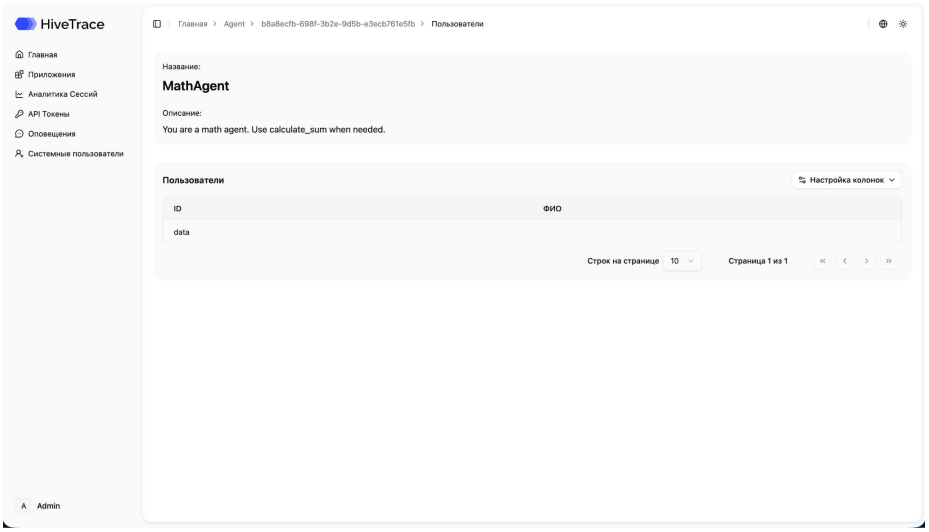

Рис. 37 - Профиль агента с вкладкой «Пользователи»

9.7. Обновление данных в реальном времени

Ключевые таблицы платформы обновляются автоматически без перезагрузки страницы благодаря технологии Server-Sent Events (SSE). Это касается: таблицы сессий, приложений, оповещений, системных пользователей, пользователей приложений, Custom Policies и счётчиков дашборда. Новые записи появляются в таблицах мгновенно.

10. Интеграция

HiveTrace может быть интегрирован в IT-инфраструктуру тремя способами.

10.1. API

HiveTrace предоставляет серверные API для мониторинга, валидации и модерации взаимодействий с LLM.

10.1.1. Аутентификация

Все запросы требуют заголовок:

```
Authorization: Bearer <API_TOKEN>
```

При отсутствии или неверном токене возвращается ответ 401 Unauthorized.

10.1.2. Base API

Два основных эндпоинта:

POST /process_request/ - анализ сообщений пользователя до их передачи в LLM.

POST /process_response/ - анализ ответов LLM перед возвратом пользователю.

Обязательные поля запроса: application_id (UUID) и message (текст). Опционально: additional_parameters с user_id и session_id.

10.1.3. Override API

Эндпоинты /process_request/override/ и /process_response/override/ позволяют задавать параметры валидации inline, игнорируя настройки Web UI. Валидация полностью управляется через поле validation_config.

10.2. SDK (Python)

SDK обеспечивает наиболее глубокую интеграцию на уровне бизнес-логики приложения.

10.2.1. Установка и конфигурация

```
pip install hivetrace[base]
```

Переменные окружения: HIVETRACE_URL, HIVETRACE_ACCESS_TOKEN, HIVETRACE_APP_ID.

Доступны синхронный (SyncHivetraceSDK) и асинхронный (AsyncHivetraceSDK) клиенты.

10.2.2. Рекомендованный поток

5. Проверить input: client.input()
6. Принять решение: можно ли вызывать LLM
7. Вызвать LLM (OpenAI, локальная модель и т.д.)
8. Проверить output: client.output()
9. Принять решение: что вернуть пользователю

10.2.3. Мультиагентные системы

SDK поддерживает интеграцию с фреймворками: CrewAI, LangChain и OpenAI Agents. Подробные инструкции доступны в соответствующих разделах документации на сайте.

10.3. Gateway (прокси-шлюз)

HiveTrace Gateway (LiteLLM + HiveTrace) выступает прокси-уровнем между приложением и LLM-провайдером. Предоставляет OpenAI-совместимый API, поэтому интеграция требует минимальных изменений клиентского кода.

10.3.1. Архитектура

Типовой поток: Application → Gateway (LiteLLM) → LLM. HiveTrace анализирует входные и выходные данные параллельно.

10.3.2. Требования к интеграции

Приложение должно:

5. Использовать base_url уровня /v1 (например: http://localhost:4100/v1)
6. Передавать Authorization: Bearer <LITELLM_MASTER_KEY>
7. Передавать заголовок X-Application-Id (UUID приложения)
8. Передавать идентификаторы пользователя и сессии

11. Устранение неполадок

Проблема	Возможная причина	Решение
401 Unauthorized при API-запросе	Токен отсутствует или неверный	Проверьте заголовок Authorization. Создайте новый токен в разделе API-токены
Приложение не появляется в списке	Не завершена регистрация	Убедитесь, что все обязательные поля заполнены и нажата кнопка сохранения
Guardrail не срабатывает	Политика отключена для данного направления	Проверьте, что Guardrail включён для input и/или output в настройках приложения
Кастомная политика блокирует релевантные запросы	Слишком узкий список тем	Расширьте и переформулируйте темы в промпте. Используйте окно тестирования
Оповещения не приходят	Не настроен канал доставки	Настройте Email или Telegram в разделе Конфигурация оповещений
Пользователь не появляется автоматически	Не передан user_id	Убедитесь, что user_id указан в additional_parameters или HTTP-заголовках
Маскирование данных не работает	Паттерн не включён или выбран тип Детекция	Проверьте настройки паттернов и выберите тип обработки Маскирование или Удаление
Gateway возвращает ошибку	Неверный base_url или отсутствует X-Application-Id	Проверьте URL (/v1) и наличие обязательных заголовков
Граф трассировки пустой	Не выбран разговор или нет мультиагентных данных	Выберите разговор в левой панели. Убедитесь, что приложение использует мультиагентный фреймворк
Blacklist не срабатывает	Паттерн не соответствует тексту	Проверьте регулярное выражение. Убедитесь, что тип паттерна (текст/файлы) совпадает с направлением проверки
Интерфейс недоступен после входа	Лицензия не активирована	Нажмите «Ввести ключ лицензии» и введите полученный лицензионный ключ
Сообщения перестали обрабатываться	Превышен дневной лимит или истекла лицензия	Проверьте окно «Лицензия и лимиты». При превышении лимита дождитесь следующего дня. При истечении лицензии - активируйте новый ключ

12. Часто задаваемые вопросы (FAQ)

Какие модели поддерживает HiveTrace?

HiveTrace поддерживает работу с любыми облачными и локальными моделями, подключаемыми через API. Зависимости от конкретного поставщика нет.

Можно ли использовать HiveTrace только для мониторинга, без блокировки?

Да. Используйте асинхронный режим приложения или тип обработки «Детекция» для персональных данных.

Как быстро подключить существующее приложение?

Самый быстрый способ - через Gateway (прокси). Достаточно изменить base_url и добавить заголовки авторизации.

Поддерживается ли интеграция с SIEM?

Да, через формат Syslog. Настройка выполняется на этапе развертывания.

Что происходит при срабатывании кастомной политики?

Система выставляет флаг custom_flagged=true. В синхронном режиме запрос может быть заблокирован с возвратом настроенного ответа.

Можно ли добавить собственные паттерны для очистки данных?

Да. Вы можете настроить пользовательские паттерны на основе регулярных выражений через интерфейс.

Какие роли доступны для системных пользователей?

Администратор (ADMIN) - полный доступ ко всем разделам. Разработчик (DEVELOPER) - приложения, политики, аналитика, API-токены; недоступны оповещения и системные пользователи.

Файлы пользователей проверяются автоматически?

Нет. Содержимое файлов не проходит автоматическую цензуру и доступно только для ручного просмотра.

Чем отличается синхронный режим от асинхронного?

В синхронном режиме проверка выполняется до передачи запроса в модель - при нарушении запрос блокируется и пользователю возвращается настроенное сообщение. В асинхронном режиме анализ происходит в фоне, без блокировки.

Как работает трассировка агентов?

На вкладке «Трассировка» в карточке приложения отображается интерактивный граф взаимодействий между агентами. Выберите разговор в левой панели - в основной области появится визуализация цепочки вызовов.

Что такое чёрные списки (Blacklist)?

Это списки запрещённых паттернов (регулярные выражения), при совпадении с которыми сообщение автоматически помечается как нарушение. Настраиваются отдельно для сообщений и файлов с указанием направления (input/output).

Где посмотреть историю операций очистки данных?

В разделе «История очисток» - там фиксируются все операции с датой, количеством удалённых сообщений/ответов и статусом.

Обновляются ли данные в таблицах автоматически?

Да. Ключевые таблицы (сессии, оповещения, приложения и др.) обновляются в реальном времени благодаря технологии SSE - без перезагрузки страницы.

Что происходит при превышении дневного лимита сообщений?

Новые сообщения временно не обрабатываются до начала следующего дня. Интерфейс, отчёты и администрирование продолжают работать без ограничений.

Что делать, если лицензия истекла?

Обработка сообщений останавливается, но функции управления остаются доступными. Получите новый лицензионный ключ и активируйте его через окно «Лицензия и лимиты».

13. Глоссарий

Термин	Определение
GenAI	Генеративный искусственный интеллект
LLM	Large Language Model - большая языковая модель
Prompt-инъекция	Техника внедрения вредоносных инструкций в запрос к модели
Jailbreak	Попытка обхода ограничений модели
PII	Personally Identifiable Information - персональные данные
Guardrail	Встроенный набор правил анализа и классификации HiveTrace
Dataclean	Модуль обнаружения признаков наличия чувствительной информации
SDK	Software Development Kit - набор средств разработки
Gateway	Прокси-шлюз для маршрутизации запросов к моделям
SIEM	Security Information and Event Management - система управления событиями
Токен	Единица текста, используемая моделью при обработке (слово или часть слова)
Сессия	Последовательность взаимодействий пользователя с AI-приложением
Red Teaming	Метод тестирования путём имитации сценариев воздействия
On-premise	Развертывание системы на собственной инфраструктуре организации
API-токен	Ключ авторизации для программного доступа к платформе
Blacklist	Чёрный список - набор запрещённых паттернов для автоматической блокировки
SSE	Server-Sent Events - технология обновления данных в реальном времени без перезагрузки страницы
Трассировка	Визуализация цепочки взаимодействий между агентами в виде графа
Лицензия	Ключ активации, определяющий срок действия и лимиты использования платформы