Руководство по развертыванию

Интеллектуальная система управления корпоративным казначейством FinAxio. Модуль «Искусственный интеллект — Среда аналитики».

1. Назначение и состав

Модуль работает внутри FinAxio. Пользователь открывает его из меню «Среда аналитики»; запросы экрана проходят через BFF по маршрутам /api/ai/*. Расчёты выполняет отдельный сервис моделей analytics-ml на порту 8091. Для обычного запуска нужны действующий контур FinAxio с PostgreSQL, Redis и BFF, миграции модуля, сервис моделей и назначенные права.

2. Порядок развёртывания

1. Запустите PostgreSQL, Redis и BFF в том же порядке и окружении, что и остальные компоненты FinAxio.
2. Примените миграции BFF командой `npm run bff:migrate`. В базе должны появиться таблицы ai_analytics_sessions, ai_analytics_trace, ai_analytics_recommendations, ai_analytics_shap_pairs и ai_analytics_inbox.
3. Запустите сервис моделей командой `docker compose up -d analytics-ml`. Образ собирается из server/src/analytics/ml; опубликованный порт привязан только к `http://[localhost]:[порт]`. Не открывайте его во внешнюю сеть.
4. Проверьте `GET http://[localhost]:[порт, по умолчанию - 8091]/health`. Код 200 и поле «ok»: true означают, что файлы моделей прочитаны и сервис готов. Код 503 означает, что процесс запущен, но не готов к расчётам; причины перечислены в поле «problems».
5. Убедитесь, что BFF обращается к сервису моделей. Адрес по умолчанию — `http://[localhost]:[порт]`; иной доступный BFF адрес задаётся переменной ANALYTICS_ML_URL. Если BFF и сервис работают в разных контейнерах, укажите адрес сервиса, достижимый из контейнера BFF.
6. Назначьте пользователю права на раздел и работу с ним, затем откройте «Среду аналитики» под его учётной записью. Состав прав приведён в руководстве администратора.

3. Настройка текстовой модели

Текстовая модель диалога настраивается отдельно от расчётного сервиса в runtime-настройках BFF: runtime.ai.gateway_api_key, runtime.ai.gateway_base_url и runtime.ai.gateway_model. В контуре без внешнего API в те же поля записывают адрес, ключ и имя модели собственного OpenAI-совместимого сервера. При недоступности текстовой модели формальный разбор фразы, чтение платежей и расчёты продолжают работать.

OpenSearch и MLflow в docker-compose.yml относятся к профилю analytics и для обычной работы экрана не требуются. Реестр MLflow подключается только при заданном MLFLOW_TRACKING_URI.

4. Проверка после запуска

Войдите под пользователем с правами просмотра и использования модуля, задайте вопрос «Платежи за март 2026 в USD» и проверьте ответ в пределах доступных ему данных. После сохранения диалога проверьте запись в ai_analytics_sessions и след в ai_analytics_trace с тем же correlation_id. Состояние сервиса моделей дополнительно подтвердите запросом /health с кодом 200.