

Руководство эксплуатации

YasnoCore LLM 1.0.0 | ООО «ЯСНОКОР» | редакция 19.10.2026

Локальный API

Метод и маршрут	Назначение
GET /health	Состояние ядра, версия и доступность бэкенда
GET /v1/models	Идентификаторы моделей от бэкенда
POST /v1/chat/completions	Передача проверенных сообщений локальной модели
POST /v1/yasnocore/analyze	Формирование задания анализа с JSON-ответом

Для chat передайте объект messages со списком ролей system/user/assistant и непустым текстом; поддерживаемые параметры генерации передаются бэкенду. Шлюз использует настроенный alias модели и выключает потоковый режим. Некорректный JSON или роль отклоняются с HTTP 400; недоступный бэкенд возвращает 503.

В analyze передаётся text и дополнительные поля, поддерживаемые build_analysis_messages. Версия 1.0.0 формирует собственный запрос со stream=false и JSON response_format; произвольные параметры генерации из входного analyze не следует считать автоматически переданными модели. Для короткой ограниченной генерации используйте chat с max_tokens.

Журналирование

Для POST сохраняются request_id, маршрут, HTTP-статус, длительность, SHA-256 входного объекта и локальный адрес клиента. Сам входной текст штатный аудит ядра не сохраняет. Веб-кабинет и inference-бэкенд могут иметь отдельные журналы; их правила хранения устанавливает администратор.

Граница доступа

Оператор работает на сервере либо через отдельный согласованный контур. Ядро 1.0.0 намеренно ограничено loopback; пользовательские роли веб-кабинета не являются авторизацией API. Не открывайте 8080/8090 наружу.

Утилиты и дополнительный кабинет

YasnoCore LLM 1.0.0 | ООО «ЯСНОКОР» | редакция 19.10.2026

Утилиты

doctor проверяет бэкенд и модели. **prepare-dataset** создаёт train/validation и manifest с контрольными суммами. **evaluate** читает JSONL с expected/actual и считает метрики. **manifest** формирует опись релиза. **smoke-train** выполняет синтетическое упражнение LoRA; **train-adapter** требует отдельного окружения и checkpoint.

Каждая строка входного набора должна содержать instruction, response, language, source_id и rights_basis. Проверьте права, происхождение и группировку источников до подготовки. Значения прав unknown/неизвестно/нет и дублирующиеся пары отклоняются. Укажите отдельный выходной каталог, сохраняйте оригинал.

Кабинет подготовки корпуса

На <https://llm.yasnosoft.ru/?page=login> партнёр входит по индивидуальному доступу. Партнёр добавляет текст и перевод; языковой эксперт проверяет содержание; администратор отдельно проверяет основание прав и допуск к выгрузке. Веб-часть использует свою MariaDB и серверные сессии. Материалы не становятся публичными от их загрузки.

Первый вход может требовать смены временного пароля. После редактирования материала допуски необходимо получить заново. Подготовка/выгрузка корпуса сама по себе не запускает обучение модели. Аудио, видео и PDF не обрабатываются Python-ядром LLM как распознавание речи.

Неисправности и поддержка

HTTP 400: проверьте JSON/поли/лимиты; 503: проверьте локальный inference; нет подключения: systemctl status и journalctl -u yasnocore-llm. Веб-кабинет при отсутствии входа переводит на форму авторизации. Не отключайте защиту и не публикуйте порт для диагностики.

Поддержка: info@yasnocore.ru, +7 (927) 841-00-61. Сообщите версию, время, команду и обезличенный фрагмент ошибки. Не передавайте пароли, корпуса и рабочие документы в обычное обращение.