

Audion Office OCR AI - руководство пользователя

Содержание

- [Быстрый путь](#)
- [Окно](#)
- [Источник и назначение](#)
- [Разделы](#)
- [Готовые файлы](#)
- [Очистка скана](#)
- [Модели и ключи](#)
- [Gemini Standard, Flex и Batch](#)
- [Основные папки](#)
- [Консольная команда](#)
- [Перед большими прогонами](#)
- [Приватность](#)

Audion Office OCR AI превращает офисные файлы, PDF, изображения и сканы в независимую от поставщика **DocumentModel** и сразу записывает выбранные офисные, архивные и машинные результаты. Markdown - необязательный экспорт для проверки и LLM, а не центр системы.

Быстрый путь

1. Запустите **Start.exe**.
2. Положите файлы в **input** или выберите другую папку или один файл как **ИСТОЧНИК** в правой колонке.
3. Откройте раздел в ленте под шапкой, затем его страницу, затем переключатель движка или поставщика.
4. Отметьте готовые файлы и нажмите кнопку запуска справа в шапке страницы.

5. Следите за терминалом; результаты - в **output** или в папке, выбранной как **НАЗНАЧЕНИЕ**.

Окно

Шапка называет машину и то, что у неё есть для OCR: Windows, компьютер, процессор, видеокарту с драйвером, память, диски, рекомендацию по видеокарте, Tesseract с языками, сервис разметки Surya и поставщиков, у которых есть ключ. Каждый бейдж говорит всё целиком во всплывающей подсказке. **RU** / **EN** переключают язык, солнце - тему. **SETUP** открывает установку того, что нужно машине для локального OCR (Tesseract, Real-ESRGAN, среды llama.cpp и Surya; шапка рекомендует нужную для этой видеокарты), - с неё начинают: пока на машине нет Tesseract, окно при запуске открывается на **SETUP**. **SETTINGS** задаёт шрифты и их размеры, **ABOUT** называет версию и папки.

Под шапкой разделы стоят одной лентой: **ЛОКАЛЬНЫЙ OCR**, **API OCR**, **ПРОВЕРКА OCR**, **ТАБЛИЦЫ**, **ИНСТРУМЕНТЫ**. Раздел с несколькими страницами показывает их второй лентой. У страницы OCR есть третья лента - движки **ЛОКАЛЬНОГО OCR**, поставщики **API OCR**; включённый переключатель закрепляет движок формы.

Форма стоит блоками с названиями; цветная полоса слева у блока говорит, какие в нём настройки. Редко нужные настройки стоят под **ДОПОЛНИТЕЛЬНО** и закрыты, пока их не откроют. Стрелка справа в шапке страницы возвращает поля страницы к значениям программы. Переключатель рискованного выбора - красный.

В правой колонке - рабочая область (**ИСТОЧНИК** и **НАЗНАЧЕНИЕ**) и терминал. Терминал показывает строку, с которой запущена каждая операция, что операция говорит своими цветами, и одну живую строку того, что она делает сейчас. **СТОП** просит операцию остановиться в ближайшей безопасной точке; повторное нажатие через несколько секунд завершает её сразу. Закрытие окна во время операции сначала спрашивает.

Источник и назначение

- **ИСТОЧНИК** - папка или один файл. **input** программы читается как есть. Источник вне программы перед запуском копируется в **input**, а **input** и **output** сначала очищаются; если в них что-то лежит, окно спрашивает перед запуском и называет, сколько. Источник внутри **output** и папка внутри **input** отклоняются до запуска.
- **НАЗНАЧЕНИЕ** - папка, в которую после запуска копируется содержимое **output**.
- Папку или файл можно бросить в окно - они станут источником. Кнопки с часами открывают историю путей, кнопка-булавка закрепляет путь сверху, **СБРОСИТЬ** возвращает оба пути к

`input\` и `output\`.

- **СПИСОК** называет в терминале, что лежит в источнике.
- Кнопка-корзина открывает очистку: содержимое источника, назначения или обоих - в Корзину, сами папки остаются; про папку вне программы окно спрашивает ещё раз. В том же меню - очистка артефактов DocumentModel.
- Кнопка со стрелками копирует Markdown из `output\` обратно в `input\`.

Разделы

ИНСТРУМЕНТЫ -> **В Markdown** извлекает выделяемый текст из DOCX, XLSX, PPTX, PDF, TXT, CSV и HTML без OCR.

ЛОКАЛЬНЫЙ OCR - бесплатный локальный OCR. Tesseract - рекомендуемая основа. Surya - необязательный медленный локальный движок с раскладкой страницы; под его переключателем строка говорит, работает ли его сервис разметки, и в ней кнопка запуска и остановки. Сервис держит модель разметки в памяти между запусками.

API OCR - OCR- и vision-модели Yandex, Mistral OCR, xAI, Gemini и OpenAI. Переключатель поставщика показывает его поля: модель, ключ, параметры. У Mistral есть управляемый второй проход: **Нет**, **Tesseract**, **Yandex** или **Yandex + Tesseract**. Gemini работает в режиме **Standard** или **Flex**; Flex дешевле, но может дольше ждать свободной мощности.

ПРОВЕРКА OCR - **Сравнение движков** (одна страница несколькими движками, сырой и очищенный OCR рядом, до большой пачки) и **Проверка чисел** (числовые и правовые строки одной страницы перечитываются vision-моделью и сверяются).

ТАБЛИЦЫ - три операции, у каждой свои настройки и своя кнопка: координатные таблицы скана в XLSX, проверка Markdown-таблиц, Markdown-таблицы в XLSX. Для линованных сканов OCR ещё восстанавливает физическую сетку, проверяет число столбцов и сверяет текст поставщика внутри подтверждённых ячеек.

ИНСТРУМЕНТЫ - **В Markdown** (быстрый перевод офисных файлов без OCR), **Проверки** (проект, окружение, модели и ключи поставщиков, короткие проверки OCR), **Workbench** (локальное ревью слабого OCR, AI-решатель сомнительных фрагментов, координаты), **Пересборка Markdown** (готовый Markdown в DOCX, LLM-DOCX, PDF, PPTX - совместимость, а не основной путь OCR) и **Gemini Batch**.

Готовые файлы

Обе страницы OCR первым показывают блок **Готовые файлы**. Рабочее значение - DOCX. Четыре офисных формата (**DOCX**, **XLSX**, **Searchable PDF**, **ODT**) и четыре формата для проверки (**Markdown**, **OCR JSON**, **HTML**, **Проверка**) независимы и делаются вместе. Каждый выбранный результат пишется из сохранённой **DocumentModel** без повторного OCR.

Пакет **<имя>.document** - каноническая внутренняя запись OCR. В нём источник, изображения страниц, кандидаты OCR, координаты, таблицы, порядок чтения, уверенность, сверка и ручные исправления. Полнота - это сохранность исходных данных, а не обещание безошибочного распознавания.

Блок **Контракт** предлагает **Текст + боксы** и **Раскладка**. Он задаёт полноту данных OCR, а не итоговый офисный формат.

Очистка скана

Блок **Очистка скана** - одна строка профилей:

- **Авто** - настроенные значения для выбранного движка.
- **Без очистки** - страница уходит как есть.
- **Тяжёлый скан** - сильнее шумоподавление, контраст и резкость.
- **Цифры** - настроено для правовых и финансовых реквизитов.
- **Ручной** - открывает подробные настройки под **ДОПОЛНИТЕЛЬНО**.

Сильная предобработка может стереть пунктуацию, тонкие линии таблиц, печати и мелкие реквизиты. Сохраняйте исходную страницу и сравнивайте очищенный образец в полном разрешении.

Модели и ключи

Список моделей открывается под своим полем. Программа хранит прочитанные списки; **ОБНОВИТЬ СПИСОК** заново спрашивает поставщика по ключу.

Ключи API лежат в **config**. Кнопка **+** у поля ключа добавляет ключ: его вставляют в окно, программа записывает новый файл в **config**, и он становится ключом поля; окно и журнал его не показывают. Кнопка-корзина рядом убирает выбранный файл ключа в **config\deleted_keys**, откуда его можно вернуть. Не кладите настоящие ключи в публичные архивы.

Gemini Standard, Flex и Batch

Standard - обычный интерактивный запрос Gemini. Для проверки одной страницы, сравнения движков и ручных опытов.

Flex - интерактивный запрос с `service_tier=flex`. Дешевле, но может дольше ждать свободной мощности. В форме это **Тариф Gemini**; у Flex более долгий таймаут.

Batch - асинхронный режим для больших несрочных прогонов. **Gemini Batch** по умолчанию готовит `workspace\gemini_batch\...\requests.jsonl` и `manifest.json`; платная отправка в Google - только с включённым **Отправить в Google**. Операция сбора проверяет задание и забирает Markdown. Подробности - в `docs\tools\GEMINI_BATCH_FLEX.md`.

Для буквального OCR через xAI сначала пробуйте `4.20 fast` (`grok-4.20-non-reasoning-latest`). `4.20 quality` (`grok-4.20-reasoning-latest`) - медленнее, для трудных таблиц и раздробленных сканов. **Проверка русского OCR** по умолчанию выключена; включённая, она даёт reasoning-модели только сомнительные страницы.

Основные папки

input\	исходные документы
output\	готовые результаты
config\	настройки, промпты и ключи API
docs\	документация проекта
docs\PDF\	документация в PDF
logs\	журналы операций
report\	служебные отчёты
workspace\	временные рабочие файлы и проверки
data\	списки моделей, которые хранит программа
cache\	кэш OCR и предобработки
App\	окно и PowerShell программы

Консольная команда

Каждая операция окна запускается из консоли через `ocr_cli.cmd` с теми же настройками:

```
ocr_cli.cmd describe
ocr_cli.cmd run --operation ocr_local_run --source D:\Scans --target D:\Results --set
ocr_engine=tesseract
ocr_cli.cmd options --operation ocr_api_run --field gemini_model --refresh
ocr_cli.cmd hardware
ocr_cli.cmd version
```

describe печатает разделы, операции и поля в JSON; **run** запускает одну операцию; **options** печатает список поля (**--refresh** спрашивает поставщика); **key add|remove** добавляет ключ из стандартного ввода или убирает файл ключа; **hardware** называет видеокарту, Tesseract и сервис разметки. Коды выхода: 0 - готово, 1 - ошибка, 130 - остановлено, 2 - неверный вызов, отклонён до запуска. Окно пишет строку запуска в свой терминал при каждом старте операции.

Перед большими прогонами

Сначала запустите **Сравнение движков** на одной странице. Для правовых документов - ещё **Проверку чисел**. Затем выберите движок, профиль очистки и готовые файлы. После OCR проверьте DOCX, XLSX и артефакты **report**; Markdown нужен, только если выбран отдельным экспортом.

Локальный OCR - когда документы должны остаться на машине, а среда поддерживает нужный язык и раскладку. Платный OCR по API - когда его возможности оправдывают внешнюю обработку и это разрешает политика данных.

Проверяйте границы строк и столбцов, объединённые ячейки, разделители чисел, даты, идентификаторы, названия организаций, подписи и печати. **Проверка чисел** - точечная сверка и не заменяет проверку всего распознанного документа.

Перед большой пачкой проверьте число источников, назначение, поставщика и модель, квоту API и свободное место на диске. Держите неудачные страницы вместе с их журналами; повторяйте только неудачные документы.

Приватность

Сканы, распознанный текст, структурные экспорты, журналы и запросы к поставщикам - чувствительные данные. В **API OCR** и AI-ресолверах страницы уходят выбранному поставщику с его ключом. Ключи API остаются в **config** и никогда не попадают в отчёты и документацию.