

Audion Office OCR AI - User Guide

Contents

- [Launch](#)
- [Root Windows](#)
- [Gemini Standard, Flex And Batch](#)
- [Scan Cleanup](#)
- [API Keys](#)
- [Before Large Runs](#)
- [Engine Selection](#)
- [Scan Cleanup](#)
- [Tables And Requisites](#)
- [Batch Control](#)
- [Failure Recovery](#)
- [Privacy](#)

Audion Office OCR AI converts office files, PDFs, images and scans into a provider-neutral **DocumentModel**, then immediately writes the selected office, archival and machine-readable results. Markdown is an optional audit/LLM export, not the center of the system.

Launch

```
launcher_gui.cmd
```

Main project folders:

input\	source documents
output\	user-facing deliverables
config\	settings, prompts and API keys
docs\	project documentation
docs\PDF\	PDF documentation
logs\	operation logs
report\	service reports
workspace\	temporary working files and checks
cache\	OCR/preprocess cache

For CLI/FZF use `launcher_project.cmd` or `launcher_project_ru.cmd`. They execute the same manifest/service layer as the GUI, so OCR Brick, Workbench resolvers and office builders do not have a separate CLI backend copy. The canonical Workbench labels are `Source`, `Add file...`, `Target`, `Reset`, `Delete`, and `List`.

Each `<name>.document` package is the canonical internal OCR record. It retains the source, page images, OCR candidates, coordinates, tables, reading order, confidence, verification and manual corrections. Completeness means preservation of source data, not a promise of perfect recognition accuracy.

Root Windows

`Office Text Without OCR` extracts selectable text from DOCX/XLSX/PPTX/PDF/TXT/CSV/HTML without AI OCR.

`Local OCR` runs free local OCR. Tesseract is the recommended baseline. Surya remains as an optional slow local engine when API cost matters more than elapsed time.

`Paid API OCR` runs Yandex, xAI, Mistral OCR 4, Gemini and OpenAI OCR/vision models. The engine is selected from one compact horizontal button row; the selected engine controls the provider-specific fields below it. Mistral supports a managed second pass: `None`, `Tesseract`, `Yandex`, or `Yandex + Tesseract`. Gemini can use `Standard` or `Flex`; Flex is cheaper, but may wait longer for available capacity. For large non-urgent runs, use `Gemini Batch OCR` in Project Tools: it prepares a JSONL/manifest locally, optionally submits the paid Google job, and `Забрать Gemini Batch` collects finished Markdown.

Both `Local OCR` and `Paid API OCR` show the shared `Deliverables` block immediately below the engine. DOCX is the working default. A balanced 4-by-2 grid provides four office formats (`DOCX`, `XLSX`, `Searchable PDF`, `ODT`) and four audit/developer formats (`Markdown`, `OCR JSON`, `HTML`,

Verification). Formats are independent and can be generated together. The checkbox grid has its own subdued dark container; the duplicate **Output formats** label is hidden while the explanatory hint remains. Full archive ZIP is still supported by the backend for compatibility, but is not exposed in the primary OCR GUI. Every selected result is written from the saved **DocumentModel** without rerunning OCR.

The **Output** section presents **Text + boxes** and **Layout** as two equal full-row buttons. This contract controls OCR data fidelity, not the final office format.

The **Maintenance** area includes a confirmed **Clear DocumentModel artifacts** action. It removes ***.document**, ***.document.json** and ***.verification.json** from managed output, report and workspace folders while preserving DOCX, XLSX, Markdown, PDF, images, sources and already-built full archive ZIP files.

For difficult Russian tables, Mistral's second pass compares provider text inside verified physical cells. Ambiguous geometry and numeric disagreements remain review items instead of silently rewriting the table.

For literal xAI OCR, try **4.20 fast** (**grok-4.20-non-reasoning-latest**) first. **4.20 quality** (**grok-4.20-reasoning-latest**) is the slower option for difficult tables and fragmented scans. The adapter uses regular non-streaming **chat/completions**, avoiding dependence on a fragile streaming route. Transient HTTP/network failures receive up to four attempts with backoff; responses are strictly validated as UTF-8 and successful pages enter the resumable cache. **Russian OCR review** is off by default. When enabled, the reasoning model reviews only suspicious pages; excessive rewrites are rejected and the original remains in metadata.

OCR Quality Test runs one-page raw/clean comparisons before large OCR jobs.

Requisites Check verifies exact legal/financial strings such as contract numbers, purchase IDs, dates and sums.

Tables exposes three expanded operations in one window without child navigation: coordinate-row extraction to XLSX, Markdown table inspection and XLSX build. For ruled scans, the OCR backend also recovers a physical grid, validates its column count and compares provider text inside verified cells.

DEV Markdown PDF renders dark/light PDFs from Markdown through Chromium. **docs\PDF** is the default. **PDF next to MD** remains available for one PDF subfolder per Markdown folder, while Mirror to Target creates a separate generated tree.

Project Tools installs and checks Tesseract, Real-ESRGAN, Surya/llama.cpp, API model lists, Gemini Batch, smoke tests and project status. It also contains the expert **Rebuild Legacy Markdown**

compatibility group for archived DOCX/PDF/PPTX/XLSX workflows; this is not the primary OCR path. The compact hardware badge appears only here and in `Local OCR`.

Detailed Gemini Standard/Flex/Batch notes and the verified dry-run are in

`docs\GEMINI_BATCH_FLEX.md`.

Gemini Standard, Flex And Batch

`Standard` is the normal interactive Gemini request. Use it for one-page checks, OCR quality tests and manual experiments.

`Flex` is an interactive request with `service_tier=flex`. It is cheaper, but may wait longer for available capacity. In the GUI this is the `Gemini tier` field; Flex uses a longer timeout.

`Batch` is the asynchronous mode for large non-urgent OCR runs. `Gemini Batch OCR` prepares `workspace\gemini_batch\...\requests.jsonl` and `manifest.json` by default; the paid Google submission happens only when `Submit to Google` is enabled. `Забрать Gemini Batch` checks the job and downloads Markdown.

Verified dry-run: `input\Бонпок 239.pdf`, pages `6-8`, `Raw`, `DPI 300`, `gemini-3.5-flash`, no Google submission. It produced 3 JSONL requests with keys `239_page_0006`, `239_page_0007`, `239_page_0008`.

Scan Cleanup

The `Preprocessing` section exposes one row of cleanup profiles. The duplicate `Scan cleanup` field label is hidden while the explanatory hint remains:

- `Auto` uses tuned defaults for the selected engine.
- `Raw` sends the page unchanged.
- `Heavy scan` boosts denoise, contrast and sharpening.
- `Numbers` is tuned for legal/financial identifiers.
- `Manual` opens the detailed preprocessing controls in Advanced.

API Keys

API keys live in `config\`. Add/delete them from the GUI next to the provider key field. Deleting a key should show a warning. Do not include real keys in public archives.

Before Large Runs

Run **OCR Quality Test** on one page first. For legal documents, also run **Requisites Check**. Then choose the engine, cleanup profile and required deliverables. Review DOCX/XLSX and **report** artifacts after OCR; Markdown is needed only when selected as a separate audit export.

Engine Selection

Use local OCR when documents can remain fully offline and the installed runtime supports the required language and layout. Use a paid OCR API when its document/layout capabilities justify external processing and the data policy permits upload. Gemini Standard, Flex, and Batch differ in scheduling, latency, and job handling; choose the documented mode rather than treating them as interchangeable.

Scan Cleanup

Apply deskew, rotation, noise removal, contrast, crop, or background cleanup only as needed. Aggressive preprocessing can erase punctuation, thin table lines, stamps, or small requisites. Keep the original page and compare a cleaned sample at full resolution.

Tables And Requisites

Check row/column boundaries, merged cells, numeric separators, dates, identifiers, organization names, and signatures/stamps. Requisites Check is a targeted validation step and does not replace review of the full recognized document.

Batch Control

Before a large batch, verify source count, output/report folders, provider/model, API quota, retry policy, and available disk space. Prevent overlapping runs against the same workspace. Follow per-document markers and reports instead of assuming that silent provider wait is a hang.

Failure Recovery

Keep failed source pages and their logs together. Retry only failed documents when the pipeline supports stable per-document output. Do not merge partial batches by filename alone if duplicate names can occur.

Privacy

Treat scans, recognized text, structured exports, logs, and provider requests as sensitive. API keys stay in `config\api_key_*.txt` and are never copied into reports or documentation. Public archives sanitize keys in isolated staging.

During release review, compare the GUI manifest with the rendered controls, command/request preview, reports, and this guide. New OCR options must explain what they change in recognition or output, not merely repeat the field label.