

Gemini OCR: Standard, Flex and Batch

Эта страница описывает, как в Audion Office OCR AI использовать Gemini для OCR после добавления **Gemini tier**, **Gemini Batch OCR** и **Забрать Gemini Batch**.

Официальные страницы Google:

- [Flex inference](#)
- [Batch API](#)

Коротко

Standard - обычный интерактивный Gemini API. Используйте его для одиночных страниц, проверки качества OCR и задач, где важен быстрый ответ.

Flex - тот же интерактивный запрос, но с **service_tier=flex**. Он дешевле, зато может дольше ждать свободную мощность или отказать при перегрузке. В проекте для Flex увеличен таймаут, потому что очередь может ждать минуты.

Batch - отдельный асинхронный режим. Проект сначала готовит локальный JSONL и manifest, затем по переключателю отправляет paid job в Google. Готовый результат забирается отдельной операцией **Забрать Gemini Batch**.

Где в GUI

Платный API OCR:

- выберите движок **Gemini**;
- выберите модель;
- выберите ключ Gemini;
- в поле **Gemini tier** выберите **Standard** или **Flex**;
- для проблем с transport/timeout можно отключить **Gemini stream**.

Тест качества OCR:

- включите API-чекбокс;
- выберите Gemini среди API-движков;
- выберите **Gemini tier**;
- сравните raw/clean на одной странице перед массовым прогоном.

Инструменты проекта:

- **Gemini Batch OCR** - подготовить JSONL/manifest и при необходимости отправить Batch job;
- **Забрать Gemini Batch** - проверить job и скачать готовый Markdown.

Gemini Batch OCR

Поля:

- **PDF/изображение/папка** - файл, папка или относительный путь от корня проекта;
- **Страницы** - **all**, одиночная страница или диапазоны вроде **6-8,12**;
- **DPI рендера** - обычно 300;
- **Препроцессинг** - **Raw**, **Auto**, **Heavy**, **Numbers**;
- **Gemini model** - модель Gemini;
- **Gemini key** - локальный key-файл;
- **Отправить в Google** - выключено по умолчанию.

Если **Отправить в Google** выключен, операция только готовит локальные файлы и ничего не отправляет наружу. Это безопасный dry-run для проверки размера и состава Batch-запроса.

Если **Отправить в Google** включен, проект:

1. Рендерит выбранные страницы.
2. Применяет выбранный preprocess-профиль.
3. Собирает **requests.jsonl**.
4. Загружает JSONL через Gemini File API.
5. Создаёт Batch job.
6. Записывает **job_name** в manifest и **workspace\gemini_batch\latest.json**.

Где лежат файлы

Каждый запуск пишет отдельную папку:

```
workspace\gemini_batch\YYYYMMDD_HHMMSS\  
  manifest.json  
  requests.jsonl  
  status.json           после проверки job  
  results.jsonl         после скачивания результата  
  pages\*.md            Markdown по страницам  
  *_gemini_batch.md     склеенный Markdown по исходному файлу
```

Последний запуск:

```
workspace\gemini_batch\latest.json
```

`latest.json` хранит путь к manifest и последний `job_name`, чтобы [Забрать Gemini Batch](#) мог работать без ручного копирования id.

Проверенный dry-run

Проверка была выполнена без отправки в Google:

```
input\Вопрос 239.pdf  
pages: 6-8  
DPI: 300  
preprocess: Raw  
model: gemini-3.5-flash  
submit: false
```

Результат проверки:

- создан manifest;
- создан `requests.jsonl`;
- строк JSONL: 3;
- ключи запросов: `239_page_0006`, `239_page_0007`, `239_page_0008`;
- первая картинка передана как `image/png`;
- prompt, system instruction и `temperature=0.0` присутствуют.

Практические рекомендации

Для одиночных страниц и OCR-бенчмарков начинайте со **Standard**.

Для больших несрочных прогонов используйте **Batch**: он удобнее для сотен/тысяч страниц и дешевле стандартного интерактивного режима, но результат приходит позже.

Flex имеет смысл, когда нужен обычный синхронный OCR-запрос, но цена важнее стабильной задержки. Если Gemini начал долго ждать, ловить transport timeout или capacity errors, вернитесь на **Standard** либо используйте **Batch**.

Для таблиц и страниц с вертикальными полосами сначала пробуйте **Raw**. Сильная очистка может помочь слабому тексту, но иногда повреждает линии таблиц и строки рядом с полосами.

English Summary

Use **Standard** for normal interactive Gemini OCR. Use **Flex** when lower cost is more important than stable latency. Use **Gemini Batch OCR** for large non-urgent jobs: it prepares a local JSONL/manifest first, submits a paid Google Batch job only when the submit toggle is enabled, and **Забрать Gemini Batch** downloads finished Markdown later.