

Нормализация Документов По Аудиту

Дата актуализации: 2026-05-14

GUI-пути:

ИСПРАВЛЕНИЕ OPENAI
ИСПРАВЛЕНИЕ GEMINI
ИСПРАВЛЕНИЕ XAI
ИСПРАВЛЕНИЕ CLAUDE

Нормализация не вызывает LLM повторно. Она использует уже готовые audit logs:

```
logs\**\*_audit.json
```

Команда **ИСПРАВЛЕНИЕ OPENAI** берёт только audit logs с `meta.provider: "openai"`. Команда **ИСПРАВЛЕНИЕ GEMINI** берёт только audit logs с `meta.provider: "gemini"`. Команда **ИСПРАВЛЕНИЕ XAI** берёт только audit logs с `meta.provider: "xai"`. Команда **ИСПРАВЛЕНИЕ CLAUDE** берёт только audit logs с `meta.provider: "anthropic"`.

Что Куда Класть

Обычный сценарий:

1. Положить DOCX/PPTX в `input\`.
2. Запустить обычный аудит OpenAI или Gemini.
3. Проверить, что audit JSON появился в `logs\`.
4. Запустить соответствующую нормализацию.

Если аудит уже готов, достаточно положить исходный DOCX в тот же относительный путь внутри `input\`, а audit JSON - в `logs\` или вложенную папку под `logs\`.

Нормализация работает только с DOCX. PPTX остаются в аудите и отчётах, но не переписываются машинным способом.

Что	Где лежит
Исходные DOCX	input\
Готовые audit JSON	logs***__audit.json
Нормализованные DOCX-копии	output\...*__normalized.docx
DOCX/XLSX/MD/JSON отчёты нормализации	output_normalization\
Машинный patch-plan	report\document_normalization\

Что Должен Дать Аудит

Для автоматической правки в audit JSON у ошибки должны быть поля:

```
{
  "fix_mode": "safe_replace",
  "old_text": "точный фрагмент из DOCX",
  "new_text": "замена",
  "confidence": "high",
  "block_id": "docx_p_0001"
}
```

Если этих полей нет, если `confidence` не `high`, если `fix_mode` не `safe_replace`, или если `old_text` не найден в блоке ровно один раз, правка не применяется автоматически.

Старые audit JSON без этих полей остаются пригодными для отчёта, но не дают автоматических замен. Нормализатор не достраивает решения сам и не обращается к LLM.

Что Получится

Normalized DOCX-копии:

```
output\<relative_path>\<stem>__normalized.docx
```

DOCX/XLSX/MD/JSON отчёты:

```
output\_normalization\<timestamp>__normalization_openai.docx
output\_normalization\<timestamp>__normalization_gemini.docx
```

Машинный patch-plan:

```
report\document_normalization\latest_normalization_patch_plan_openai.json
report\document_normalization\latest_normalization_patch_plan_gemini.json
```

Правило Безопасности

LLM не редактирует DOCX напрямую. Аудит только записывает возможные безопасные правки в JSON. Нормализация применяет их локально и только к копиям.

Всё сомнительное остаётся в таблице **Unresolved Items** в DOCX-отчёте.

CLI Для Проверки

GUI вызывает тот же локальный скрипт:

```
runtime\python.exe system_core\document_normalizer.py --provider openai --from-logs
logs
runtime\python.exe system_core\document_normalizer.py --provider gemini --from-logs
logs
```

Для сухого прогона без записи DOCX:

```
runtime\python.exe system_core\document_normalizer.py --provider openai --from-logs
logs --dry-run
```