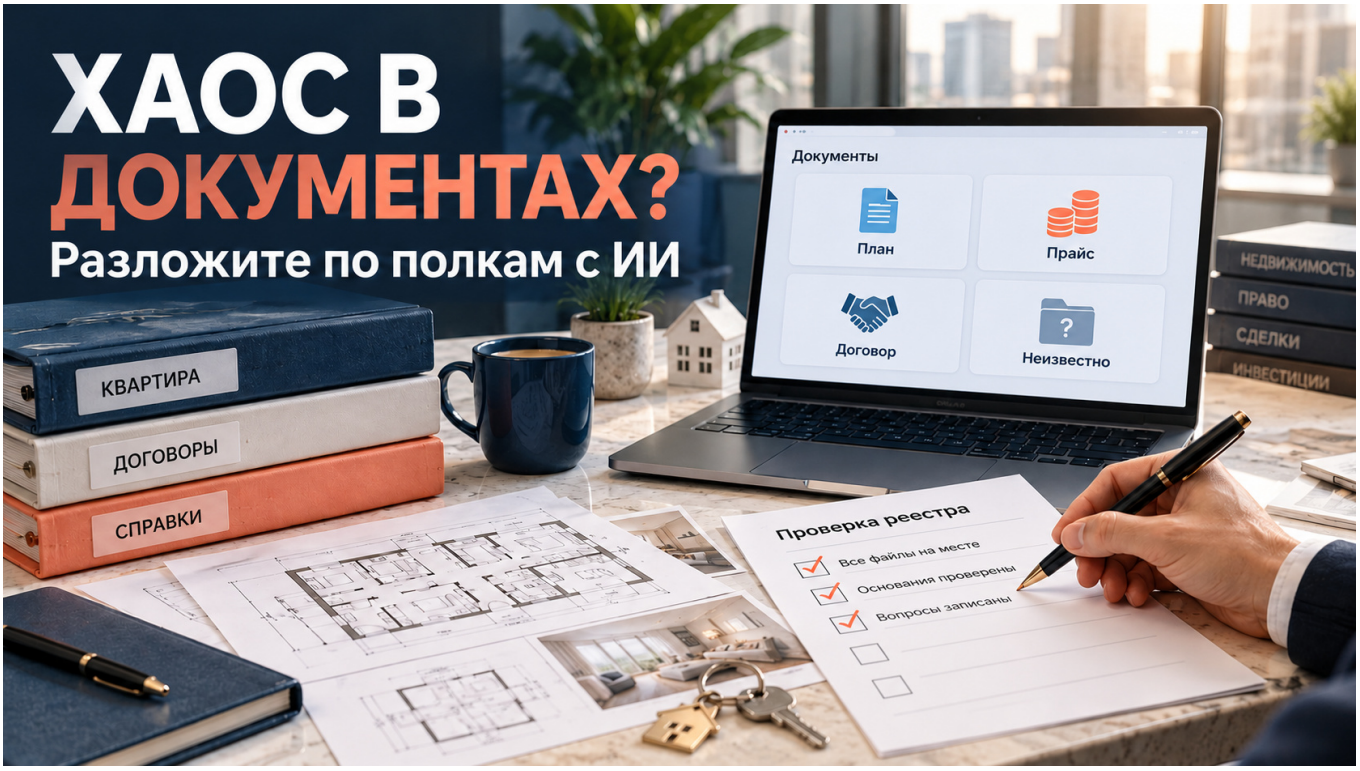
Документы по полкам: классификатор и журнал ручной проверки

Соберите реестр, в котором каждый файл можно найти, а каждый предложенный тип — проверить по содержанию.



12 страниц · 6 шагов · 18 проверок

Словарь категорий, пять размеченных примеров, инструкция для нейросети, таблица ошибок и два рабочих листа.

Авторский процесс классификации материалов по типам. Он не подтверждает юридическую силу, подлинность и актуальность документов и не оценивает людей.

С чего начать

В папке сделки смешаны планы, прайсы, договоры и непонятные сканы. Нужно подготовить прозрачный реестр для себя или передачи коллеге.

Поле	Пример	Зачем
ID	A-01	Сохранить связь с исходником
Тип	План	Найти нужный вид материала
Основание	Видны комнаты и размеры	Проверить решение по содержанию
Статус	Не проверено человеком	Не подменить экспертизу меткой

Подготовьте исходные данные

- Пять учебных или разрешённых к обработке файлов разных типов.
- Словарь категорий с положительными и отрицательными примерами.
- Исходная ссылка и постоянный ID каждого файла.
- Эталонные ответы человека, подготовленные до запуска.
- Таблица реестра без лишних персональных данных.

Соберите и сопоставьте факты

1. Выберите одно решение

Запишите, для чего нужен реестр: найти типы материалов и вопросы по конкретной папке. Отложите переименование, удаление и отправку клиенту. Укажите, кто будет проверять результат и какой набор файлов входит в запуск.

Готово: Цель описана одним предложением; исходники не меняются.

Не делайте так: Просить «разобраться с архивом» без границ и получателя результата.

2. Соберите независимые поля

Отделите тип от темы, формата файла и статуса проверки. Для типа задайте словарь: план, прайс, договор, фото объекта, переписка, другое, неизвестно. Для составного файла разрешите несколько типов.

Готово: В словаре есть определения, исключения и допустимый ответ «неизвестно».

Не делайте так: Делать «срочно», «проверено» и «ипотека» вариантами одного поля типа.

3. Подготовьте эталон

Просмотрите тестовый набор и разметьте его до запроса к модели. Включите размытый скан, составной файл и документ вне словаря. Проверьте доступность текста после распознавания и исключите лишние данные.

Готово: Для каждого ID есть эталон или честная запись, что нужен новый исходник.

Не делайте так: Проверять только одинаковые простые файлы или принимать ответ модели за эталон.

Что нужно получить перед продолжением

Примите и зафиксируйте решение

4. Запросите метку с основанием

Дайте словарь и требуемую таблицу: ID, тип, фрагмент-основание, вопрос. Запретите домысливать подлинность, актуальность и принадлежность объекта. Для нечитаемого файла требуйте вопрос, а не наиболее похожую категорию.

Готово: В ответе столько же ID, сколько во входном реестре; каждый фрагмент можно найти.

Не делайте так: Принимать проценты уверенности, написанные моделью, за доказанную точность.

5. Разберите каждое несовпадение

Сравните тип с эталоном и основание с исходником. Отдельно запишите неверный тип, пропуск ID, выдуманный фрагмент и корректный отказ от классификации. Исправьте правило или исходный текст; повторите весь тест, если словарь менялся.

Готово: Есть журнал ошибок с причиной, исправлением и повторной проверкой.

Не делайте так: Считать неизвестный ответ ошибкой и принуждать модель выбрать что-нибудь.

6. Передайте реестр с историей

Сохраните версию словаря, дату и ответственного. Передавайте коллеге ссылки на исходники и незакрытые вопросы. Новые файлы обрабатывайте тем же порядком; выдача клиенту остаётся отдельным проверяемым действием.

Готово: Коллега находит исходник и видит, что подтверждено, а что ещё требует внимания.

Не делайте так: Автоматически отправлять файл клиенту сразу после появления метки.

Что проверить после выполнения

Заполненный пример

Учебный набор. Названия файлов условные; никакие персональные данные не использованы.

ID / название	Содержимое	Тип / действие
A-01 · scan.jpg	Комнаты и размеры	План; сверить объект
A-02 · prices.pdf	Лоты, площади и цены	Прайс; уточнить дату
A-03 · final.pdf	Договор и план в приложении	Составной; отметить части
A-04 · photo.jpg	Размытый нечитабельный лист	Неизвестно; запросить копию
A-05 · payment.pdf	Получатель и сумма к оплате	Другое; решить вопрос категории

Пять исходников — пять строк. Метка «план» не подтверждает, что это актуальный план нужного объекта. Эти вопросы остаются отдельными.

Как сделать запись полезной

Общая оценка	Проверяемая запись
Есть сумма → прайс	Прайс содержит перечень предложений с ценами
Имя final → актуально	Актуальность подтверждается отдельной проверкой
Скан не читается → договора нет	Текст не читается; нужна другая копия

Выберите действие по ситуации

Наблюдение	Что делать	Что не делать
Два типа в одном файле	Метка «составной» и описание частей	Удалять «лишние» страницы
Текст не распознан	Запросить читаемый исходник	Делать вывод о содержании
Основание выдумано	Отклонить ответ и проверить инструкцию	Сохранять красивую метку
Типа нет в словаре	Ручная проверка и версия словаря	Подгонять под ближайший тип

Уточните перед решением

1. Какое решение ускоряет этот реестр?
2. Где сохранён исходник каждого ID?
3. Кто подтвердит результат и каким действием?
4. Какой ответ допустим, если данных недостаточно?

18 проверок результата

Галочка означает подтверждённый результат. Если данных нет, перенесите вопрос в лист решений.

До запуска

- [ ] **Определить задачу.** Найти типы и незакрытые вопросы в одной папке. Нет конкретного решения — сузьте цель.
- [ ] **Присвоить ID.** Один исходник соответствует одной строке. Файл без ID не включайте в пакет.
- [ ] **Сохранить исходники.** Копии и исходные ссылки доступны. Нет восстановления — не меняйте файлы.
- [ ] **Убрать лишние данные.** Учебные файлы не содержат персональных реквизитов. Неясны правила обработки — используйте вымышленные примеры.
- [ ] **Разделить поля.** Тип, тема и проверка записаны отдельно. Срочно не является типом документа.
- [ ] **Разрешить неизвестное.** Нечитаемый файл получает вопрос. Не принуждайте выбирать случайную категорию.

На тестовом наборе

- [ ] **Составить эталон.** Человек разметил файлы до запроса. Ответ ИИ не становится эталоном сам по себе.
- [ ] **Добавить сложные случаи.** Есть размытый скан, составной и незнакомый тип. Одинаковые планы не проверяют границы.
- [ ] **Сверить ID.** Количество и сами идентификаторы совпали. Пропуск строки требует исправления.
- [ ] **Проверить основания.** Каждый фрагмент найден в исходнике. Выдуманную цитату отклоните.
- [ ] **Разделить ошибки.** Неверный тип и неизвестное посчитаны отдельно. Не скрывайте отказы общей цифрой.
- [ ] **Повторить тест.** После правки словаря проверены все примеры. Один исправленный файл не подтверждает весь процесс.

Перед передачей

- [ ] **Отметить проверившего.** Есть имя и дата ручного подтверждения. Метка ИИ не равна профессиональной проверке.
- [ ] **Записать версию.** Словарь и дата запуска сохранены. Без версии трудно объяснить расхождения.
- [ ] **Назначить вопросы.** По каждому неизвестному есть действие. Пустая строка не означает готовность.
- [ ] **Открыть ссылки.** Коллега находит исходные файлы. Метка без исходника не помогает проверке.
- [ ] **Посчитать два показателя.** Точность и покрытие указаны с разными знаменателями. Не делите правильные типы на удобное число.
- [ ] **Разделить выдачу.** Отправка клиенту проходит отдельную проверку. Не связывайте сортировку с автоматической отправкой.

Заполните свой рабочий лист

Один ID — одна строка. Сохраните ссылку на исходник и точный фрагмент, на котором основана метка.

Код задачи: _____ Дата: _____

ID / ссылка	Тип и основание	Проверка / вопрос

Принятое решение: _____

Что осталось неизвестным: _____

Следующий шаг и дата: _____

Вопросы, ответственность и сроки

Опишите несовпадение без общих оценок. После изменения правила проверьте весь эталонный набор.

ID / несовпадение	Причина / исправление	Повторный тест

Пять сигналов, при которых нужно уточнение

Пропал ID. восстановите строку до любой оценки качества.

Найдена выдуманная цитата. не используйте ответ как доказательство.

Все ответы «правильные» без эталона. качество ещё не проверено.

Нечитаемый файл получил уверенный тип. проверьте правило отказа.

Классификация запускает удаление. разъедините эти действия.

Согласуйте следующий шаг

Формулировка для разговора

В реестре пять файлов. Три определены по содержанию, один составной, по одному нужна читаемая копия. Типы пока предложены, проверка актуальности отмечена отдельной колонкой. Исходники сохранены по ссылкам.

Первое действие после чтения

Возьмите пять разрешённых к обработке учебных файлов. Запишите эталон, выполните один запуск и разберите все несовпадения. Увеличивайте объём после исправления правил.

Контакт ответственного: _____

Когда вернуться к проверке: _____

Другие материалы Нейросистемы →

Какие действия отнимают время после обращения?

Диагностическая карта для риелтора поможет связать работу с материалами, скорость ответа и подготовку следующего шага клиента.

Найти узкое место в работе →

Подготовлено 03.10.2026. Все файлы и показатели ниже учебные. Проверьте условия обработки данных выбранного сервиса до работы с клиентскими документами.

NIST AI 600-1: риски генеративных моделей

Промпт для первого запуска

Вставьте свой словарь и учебные фрагменты после инструкции. Не используйте реальные секреты и персональные данные для настройки.

Классифицируй материалы только по доступному содержанию и заданному словарю. Для каждого ID верни тип, точный короткий фрагмент-основание и вопрос при нехватке данных.

Если текст не читается — неизвестно. Если в файле несколько документов — составной и список обнаруженных типов. Если категории нет — другое, требуется ручной просмотр.

Не определяй подлинность, актуальность, законность и владельца по догадке. Не переименовывай, не перемещай, не удаляй и не отправляй исходники. Верни таблицу без пропущенных ID.

Два показателя вместо красивой общей цифры

Учебный пример: 20 файлов, 18 классифицированы, из них 16 верно. Два остальных обоснованно оставлены неизвестными.

Показатель	Расчёт	Как читать
Точность среди классифицированных	16 / 18 = 88,9%	Два неверных типа нужно разобрать
Покрытие классификацией	18 / 20 = 90%	Два файла ждут ручного действия
Потерянные ID	0 из 20	Реестр сохранил все входные файлы

Эти числа описывают только учебный набор. На других файлах качество может отличаться. Отдельно контролируйте пропуски и неверные объединения — средняя доля правильных типов их скрывает.