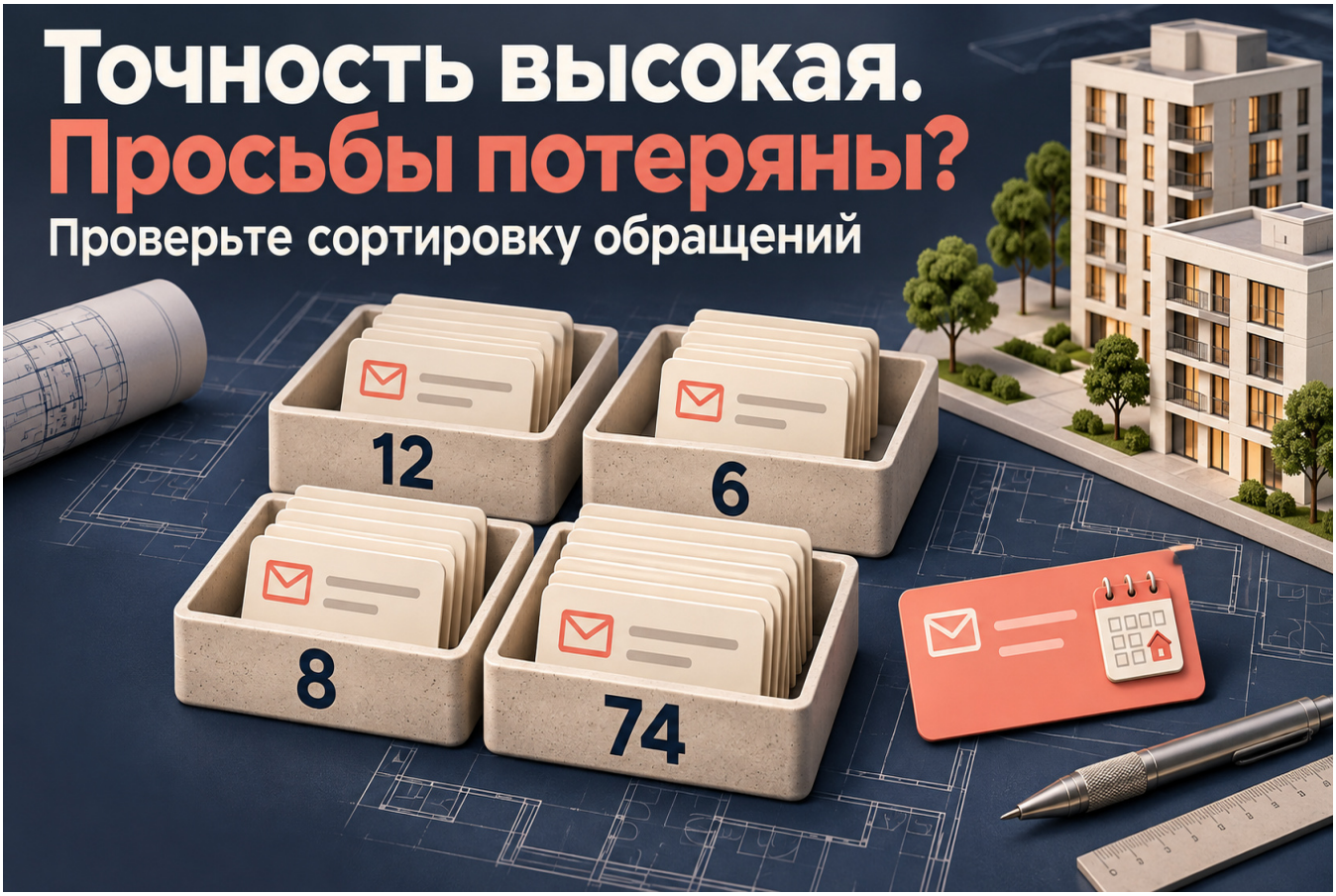
Точность высокая. Просьбы потеряны: аудит сортировки К183

Рабочая книга аудита ИИ: две матрицы, примеры обращений, расчёт precision и recall, журнал пропусков и проверка перед автоматизацией.



Практический гайд с заполненным учебным примером и чистыми рабочими листами. Подготовлено редакцией Нейросистемы риелтора.

Учебные цифры не являются данными о текущем рынке. Обложка представляет концептуальную иллюстрацию.

1. Что проверяем в обращении

K183 проверяет одну метку: есть ли явная просьба организовать или согласовать просмотр объекта. Положительный класс означает наличие признака в сообщении. Он не означает хорошего клиента, прогноз покупки или приоритет по личным характеристикам.

Поле	Учебное решение
Единица	Самостоятельное сообщение с достаточным контекстом
Да	Просьба организовать показ
Нет	Такой просьбы в данном тексте нет
Неизвестно	Недостаточный контекст или спорный случай
Действие	Подсказка агенту, без удаления сообщений

Ваша единица	Правило да	Ручная проверка

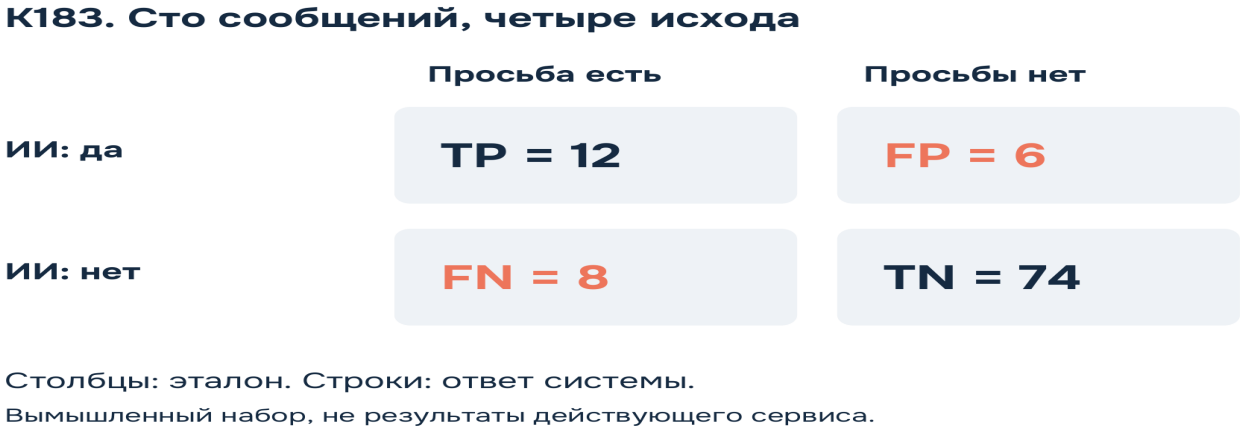
2. Примеры разметки до запуска

Сообщение	Эталон	Почему
Можно посмотреть завтра после шести?	Да	Просит согласовать просмотр
Организуйте, пожалуйста, показ этой квартиры	Да	Явная просьба
Пришлите планировку	Нет	Запрашивает файл
Хочу посмотреть схему на картинке	Нет	Просмотр изображения, не объекта
А как насчёт завтра?	Неизвестно	Нужен контекст предыдущих реплик

Эталон составляется до просмотра ответа модели. Пограничные случаи обсуждаются по письменному правилу. Сделка в будущем не меняет содержание исходной просьбы.

Спорная фраза	Уточнение правила	Решение проверяющего

3. Четыре клетки первой версии



Вымышленный набор, не результат работы действующего сервиса.

Клетка	Смысл	Число
TP	Просьба есть, найдена	12
FP	Просьбы нет, система отметила да	6
FN	Просьба есть, пропущена	8
TN	Просьбы нет, система отметила нет	74

Всего 100 размеченных сообщений. Реальных просьб 20, остальных 80. Положительных меток 18, отрицательных 82. Отдельные неизвестные сообщения в эту учебную сотню не входят.

4. Разные вопросы, разные знаменатели

Показатель	Формула в примере	Ответ
Accuracy	(12 + 74) / 100	86%
Recall	12 / (12 + 8)	60%
Precision	12 / (12 + 6)	66,7%
Пропущено реальных просьб	8 / 20	40%

Recall спрашивает: сколько реальных просьб найдено? Precision спрашивает: сколько выданных положительных меток правильные? Не называйте эти две величины одним неуточнённым словом «точность».

Что хотим узнать	Правильный знаменатель	Результат

5. Почему общего процента мало

Правило «всем нет» на этой сотне дало бы 80 правильных ответов и 80% accuracy. Но TP = 0, FN = 20: ни одна реальная просьба не найдена. Нельзя оценивать пользу для агента только общей долей совпадений.

Проверка	Решение
Метка да никому не дана	Precision имеет нулевой знаменатель
Нет реальных просьб в наборе	Recall имеет нулевой знаменатель
Показатель не определяется	Так и записать, не заменять 100%
Неизвестные сообщения	Показать количество отдельно
Сбой обработки	Не считать отрицательной меткой

Ограничение набора	Что добавить в следующую проверку

6. Сравнение двух версий

K183. Меньше пропусков, больше проверок

V1
8 пропусков
6 ложных меток
18 сообщений с меткой да

V2
3 пропуска
15 ложных меток
32 сообщения с меткой да

Один учебный набор. Нагрузка выросла на 14 проверок.
Выбор требует оценки процесса и отдельного нового набора.

Вымышленный набор, не результат работы действующего сервиса.

Показатель	V1	V2
TP / FP / FN / TN	12 / 6 / 8 / 74	17 / 15 / 3 / 65
Accuracy	86%	82%
Recall	60%	85%
Precision	66,7%	53,1%
Меток да	18	32

V2 пропускает меньше просьб, но даёт больше ложных срабатываний. Нагрузка на проверку положительных меток выросла на 14 сообщений. Это учебный компромисс, не готовый выбор для любой команды.

7. Журнал разбора ошибок

Ошибка	Учебная причина	Изменение
FN: просьба с предложением времени	Правило искало только слово показ	Добавить понятный признак согласования просмотра
FP: посмотреть план на картинке	Не различён объект действия	Уточнить, что речь о просмотре объекта
Неизвестно: как насчёт завтра	Нет предыдущих реплик	Запросить контекст, не угадывать

ID	Причина	Правка правила	Проверка

Не переводите каждый пропуск в потерянную комиссию. Записывайте наблюдаемое последствие: лишняя проверка или задержка согласования.

8. Повторная проверка без подгонки

Шаг	Что сохранить
До запуска	Версия правила и независимая разметка
При запуске	Модель или сервис, дата, разрешённые данные
При разборе	Каждая клетка и список неизвестных
После настройки	Отдельный отложенный набор
При сравнении	Одинаковая единица и явный состав наборов
В рабочем потоке	Ручной просмотр обеих меток

Нельзя настроить правило на ошибках сотни, снова проверить только эту сотню и считать доказанной работу на новых сообщениях.

Версия	Набор	Что изменилось

9. Пустой рабочий лист

ID	Эталон	Ответ системы	Основание

Клетка	Ваше число
TP	
FP	
FN	
TN	

Сначала проверьте сумму клеток и соответствие исходным ID. Затем считайте показатели. Неизвестные и необработанные записи перечислите отдельно.

10. Решение о следующем шаге

Наблюдение	Действие
Много FN	Разобрать пропущенные просьбы и формулировки
Много FP	Проверить причины ложных меток и нагрузку
Спорный эталон	Уточнить правило с проверяющим
Набор изменился	Не сравнивать проценты без описания состава
Хочется автоматического удаления	Сохранить очередь и ручной контроль

Что улучшили	Чем проверили	Какой следующий набор

Источники проверены 12.11.2026. Все числа и сообщения учебные. Матрица оценивает обнаружение текстового признака, а не людей и не вероятность покупки.

Источники и проверка

Google Developers: Thresholds and the confusion matrix

Google Developers: Accuracy, recall, precision