



Protocol: архитектура по этапам и проверка КЗ перед отправкой в Минздрав

Модуль анализа клинических протоколов и консультативных заключений - разработка МЦ «Кравира». Интеграция с МИС «Айболит» (ЕРАМ) - следующий этап.

Июнь 2026 · Версия сборки: GET /api/version · API: GET /api/version · Нормативная база: FHIR BY 5.0, Протокол МИС ОЗ-ЦИСЗ v.1.4, программа испытаний МИС v.1.3-4 (амбулаторный профиль)

Параметр	Значение
Заказчик пилота	МЦ «Кравира» (частная ОЗ, интеграция с МИС «Айболит», ЕРАМ)
Стек сервера	Python 3.11+, FastAPI (rag_server.py), детерминированные правила + RAG + LLM (опционально)
Корпус КП	~478 PDF клинических протоколов Минздрава РБ, ~55 000 фрагментов в lazy-индексе (manifest + disk)
FHIR	HL7 FHIR 5.0, профили fhir.by (пакет fhir.by#1.4.411)
Роль в контуре ЦИСЗ	Предимпортная экспертиза КЗ и Bundle до ЭЦП ; Protocol не вызывает ЦИСЗ и не подписывает пакеты

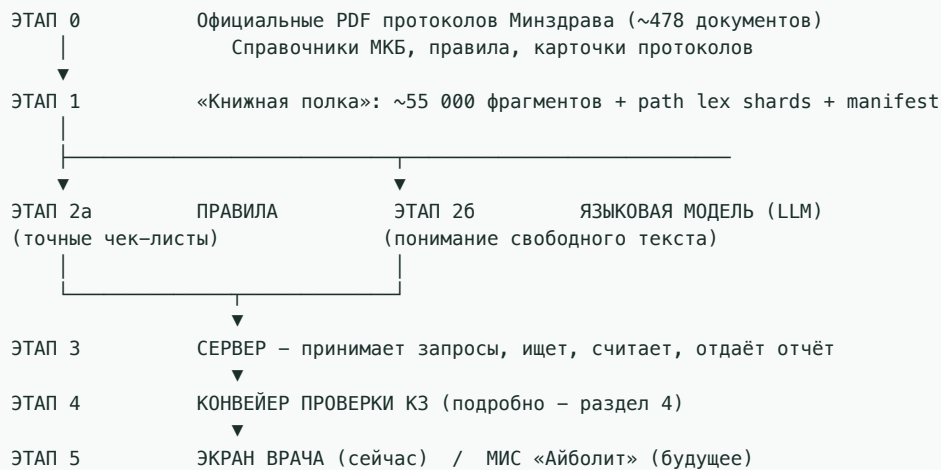
1. Задача системы

Медицинский центр обязан вести консультативные заключения по стандартам и клиническим протоколам Минздрава РБ. **Плохо оформленное или клинически слабое КЗ**, попавшее в государственный контур (ЦИСЗ), создаёт репутационный и методический риск перед Минздравом.

Protocol помогает *до подписи и отправки*:

- найти подходящий клинический протокол по диагнозу и МКБ-10;
- сравнить текст заключения с требованиями протокола;
- выставить понятную оценку по блокам и показать цитаты из официального PDF протокола;
- при необходимости - **заблокировать подпись** (политика настраивается в МИС).

2. Общая архитектура: пять этапов + два «мозга»



Правила – один и тот же документ всегда даёт один и тот же балл; в отчёте видно, *какой пункт протокола* нарушен.

LLM – помогает понять длинный текст и сформулировать итог; **итоговый процент для МИС берётся из правил**, а не «на глаз» у модели.

3. Этапы 0-3: от протоколов Минздрава до сервера

Этап 0 - корпус

Синхронизация с порталом Минздрава: PDF клинических протоколов, каталог по рубрикам, сотни формализованных правил («при диагнозе X обязательно назначить Y»).

Этап 1 - подготовка поиска

Каждый протокол режется на фрагменты (абзацы, разделы «Диагноз», «Лечение», таблицы). Строится индекс - как в картотеке библиотеки.

Этап 2 - два слоя анализа

Описано выше. Ключевое для руководства: **для блокировки отправки и отчётности Минздраву опираемся на правила**; LLM - вспомогательный слой.

Этап 3 - сервер

Программа в дата-центре: загружает корпус при старте, принимает файл или структурированные данные из «Айболит» (FHIR Bundle), возвращает JSON-отчёт с оценкой и рекомендацией `send_gate` - можно ли подписывать.

3.1. Семантический RAG: пять фаз ускорения (реализовано)

Аддитивный контур поверх `lex+BM25+rules`. **Итоговый % и `send_gate` не меняются** - фазы ускоряют поиск фрагментов и улучшают recall выдержек.

Фаза	Модули	Эффект
1	<code>build_chunk_embeddings.py</code> , <code>RAG_PRECOMPUTED_CHUNK_EMBED</code>	Офлайн-эмбединги ~55k чанков; 1 API-вызов на запрос вместо ~44
2	<code>vector_index.py</code> , <code>build_vector_index.py</code> , <code>RAG_VECTOR_INDEX</code>	Vector top-200 → lex/BM25 → embed-rerank top-44
3	<code>semantic_rule_fallback.py</code>	Синонимы для <code>required_exam</code> / <code>keyword_presence</code> после <code>substring</code>
4	<code>consult_tiering.py</code> , <code>/api/consult-review/tier</code>	L0/L1/L2 для массового потока МИС
5	<code>enrich_protocol_rules_batch.py</code>	Batch LLM → правила в <code>rule_checker</code>

Развёртывание: `corpus_pipeline` → `build_chunk_embeddings.py` → `build_vector_index.py` → `corpus_manifest.py`. Eval: `search_quality_eval.py --embed-on --precomputed-on --vector-on`.

3.5. Подбор клинических протоколов - подробно

В проекте **два независимых механизма** подбора КП. Для врача важно различать их в отчёте.

3.5.1. Карточки каталога (match_protocol_cards)

Модуль protocol_match.py сопоставляет КЗ с карточками из data/catalog/ и protocol_summaries/. Веса match_score:

Фактор	Вес	Источник
МКБ-10	40%	Блок «Диагноз», Condition.coding, словарь diagnosis_icd.py
Текст диагноза	20%	Пересечение слов с title/condition карточки
Специальность врача	15%	Рубрика из шапки КЗ → slug каталога
Аудитория (взрослый/детский)	10%	population карточки vs возраст пациента
Демография	5%	Пол, беременность
Обследования в тексте	5%	Маркеры исследований в КЗ
Жалобы	5%	Токены жалоб vs condition_label

Дополнительно: annotate_applicability() - применимость КП (амбулатория/стационар, возраст). Специальные эвристики: при I80-I89 приоритет венозных КП, штраф карточек по сердечной недостаточности при флеботромбозе.

3.5.2. RAG-отбор PDF-фрагментов (retrieve)

Модуль rag_server.retrieve() - подбор *конкретных абзацев* из корпуса. На Render поиск и КЗ ограничены allowlist путей (RAG_FORBID_FULL_CORPUS_RETRIEVE=1).

Запрос (жалобы + МКБ + фокус КЗ)

- └ [опц.] Vector index: top-200 по cosine (RAG_VECTOR_INDEX)
- └ Lex + inverted index + BM25 blend (RAG_LEX_BM25_ALPHA)
- └ Routing boost: рубрика, МКБ, возраст, matched PDF paths
- └ Embed-rerank пула ~44 чанков:
 - └ precomputed embedding в JSONL ИЛИ live Gemini embed
- └ Hybrid: $\alpha \cdot \text{lex} + (1-\alpha) \cdot \text{cosine}$ → top-N фрагментов с page/section

В режиме КЗ (CONSULT_REVIEW_STRICT_PROTOCOLS=1): RAG ограничен allowlist PDF, сопоставленных с МКБ и matched cards. Второй проход RAG - при низком max_score первой выдачи.

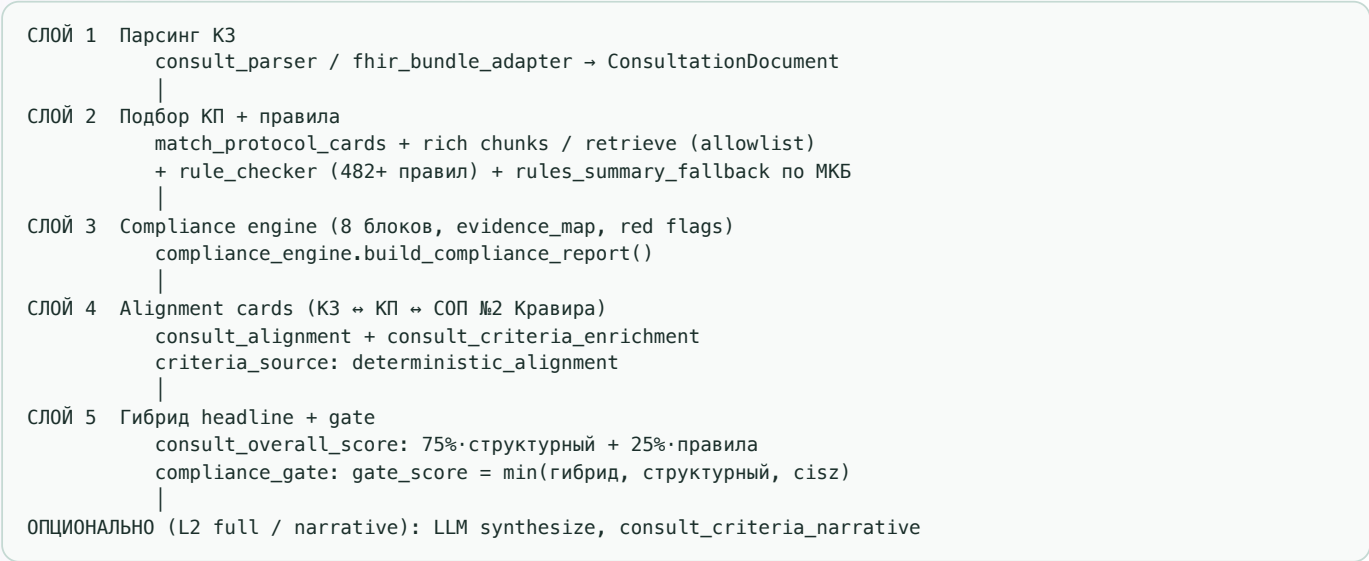
3.5.3. Различие для методслужбы

	Карточки каталога	RAG-фрагменты PDF
Что показывает	«Какие КП теоретически релевантны»	«Какие абзацы протокола цитируем»
Риск ложного матча	Широкая рубрика (I80 → кардиология)	Ниже при strict allowlist
Используется в	Блок «Применимость протокола», rule_checker scope	LLM-критерии, evidence_map, rules_compliance_pct

Пример pl_1_f: RAG верно подобрал КП по ТГВ; карточки каталога дополнительно показали КП по СН - врачу ориентир на блок «Протоколы по тексту заключения».

3.6. Архитектура оценок и анализа - пять слоёв

Итоговый балл для МИС строится **снизу вверх**. В production (июнь 2026) критерии для врача в первую очередь из **детерминированного alignment**, а не из LLM.



3.6.1. Формулы итогового процента

Показатель	Формула	Модуль	Для send_gate
Структурный %	Взвешенное среднее 8 блоков (веса из compliance_weights.yaml)	compliance_engine.py	Да
Правила протокола %	Доля passed правил из RAG-фрагментов + catalog	rule_checker.py	Да (25% гибрида)
Гибрид (headline)	0,75·структурный + 0,25·правила; затем safety cap	consult_overall_score.py	Да
LLM среднее	Среднее score_pct по критериям модели	LLM synthesize	Нет
gate_score	min(гибрид, структурный, cisz_score)	compliance_gate.py	Да
cisz_score	Чек-лист готовности FHIR Bundle (слои A-C)	cisz_readiness.py	Да (min с клиникой)

Fallback без структурного разбора: 60%·LLM-критерии + 40%·правила (CONSULT_STRUCTURED_ANALYSIS=0).

3.6.2. Типы правил (rule_checker)

rule_type	Что проверяет	Метод
diagnosis_formula	Компоненты формулировки диагноза	Маркеры + подстроки
required_exam	Обязательное обследование в K3	Substring + semantic fallback (фаза 3)
keyword_presence	Red flags, ключевые термины	Substring + алиасы (УЗИ ↔ ультразвук)
red_flag_rule	Опасные признаки протокола	Набор keywords
population_mismatch	Детский/взрослый КП vs пациент	Сравнение audience
diagnostic_criterion	Критерии нозологии	Длительность, частота симптомов

3.6.3. Уровни проверки (L0 / L1 / L2)

Уровень	API	Что включено	UI / Render
L0	/api/consult-compliance-screen, tier=L0	Парсинг + 8 блоков + send_gate + cisz; без alignment/RAG/LLM	< 2 с; 100% потока МИС
L1	/api/consult-review/tier tier=L1	Structured + alignment cards + СОП Кравира + rich chunks по matched PDF	Кнопка по умолчанию (~5-15 с)
L2	/api/consult-review, stream, tier=L2	Production: L2-lite = L1 + evidence pack + L2-fast. Local: full RAG + LLM	~15-30 с на Render

Ключевые env на Render (render.yaml): CONSULT_RENDER_L2_LITE=1, CONSULT_L2_FAST=1, CONSULT_ALIGNMENT_ENABLED=1, CONSULT_REVIEW_FORBID_FULL_CORPUS=1, CONSULT_REVIEW_CACHE=1.

3.6.4. Детерминированное согласование (alignment)

Модули consult_alignment.py, consult_criteria_enrichment.py, kravira_sop_rules.py (СОП №2, папка doc_kravira). По каждому блоку КЗ (жалобы, анамнез, статус, диагноз, обследование, лечение, наблюдение) формируется карточка:

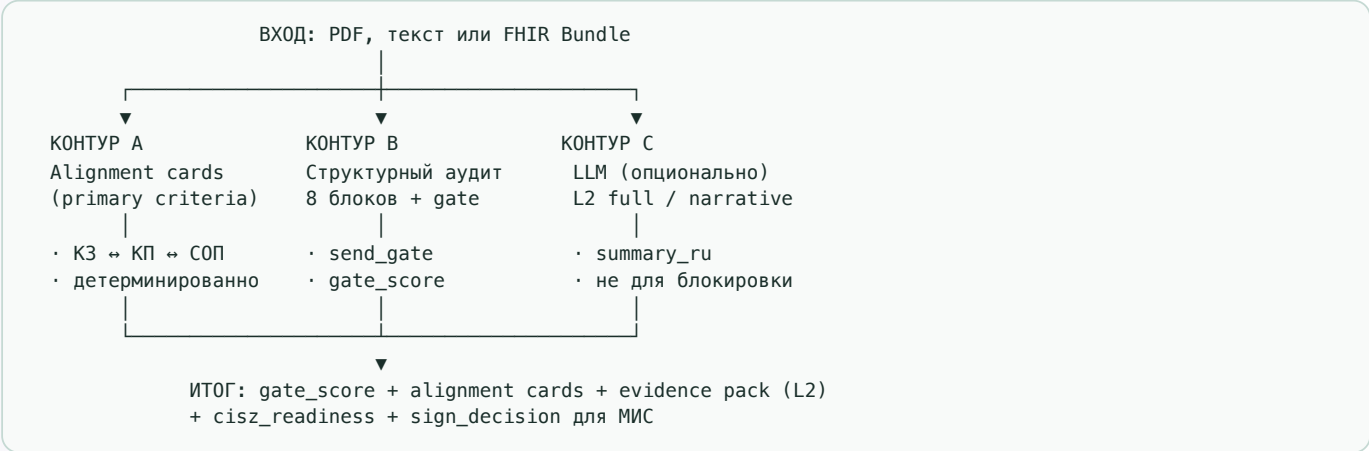
- фрагмент из КЗ и выдержка из клинического протокола Минздрава;
- эталон по СОП Кравира (оформление амбулаторной карты);
- оценка покрытия, gaps/findings, comment_ru для врача.

При CONSULT_ALIGNMENT_PRIMARY=1 (default) критерии в отчёте помечаются criteria_source: deterministic_alignment - воспроизводимый аудит для методслужбы. Опционально CONSULT_CRITERIA_NARRATIVE=1 - LLM переформулирует комментарии без новых фактов.

4. Блок анализа КЗ - подробно

Это **ядро** проекта. Ниже - что именно происходит после нажатия «Проверить» и как формируется оценка перед отправкой в Минздрав.

4.1. Три контура оценки (июнь 2026)



4.2. Семь шагов конвейера (контур В + поиск)

- 1 Чтение файла.** Из PDF извлекается текст. Скан без текстового слоя - низкое качество; в отчёте будет предупреждение и пониженная уверенность.
- 2 Разбор на «карточку заключения».** Программа ищет блоки: шапка, жалобы, анамнез, объективный статус, диагноз, назначения, рекомендации, контроль, подпись врача. Внутренний формат один - *ConsultationDocument* - независимо от того, PDF это или данные из МИС.
- 3 Диагноз и МКБ-10.** Коды из блока «Диагноз»; при отсутствии - подсказка по словарю и LLM. Специальность врача из шапки не даёт уйти в чужую рубрику протоколов.
- 4 Подбор клинических протоколов.** По МКБ, рубрике и тексту выбирается 1-6 официальных КП Минздрава.
- 5 Выдержки из протоколов.** Из PDF протоколов достаются конкретные абзацы: что рекомендовано при таком диагнозе.
- 6 Детерминированная проверка** (см. 4.3) - восемь блоков оценки, красные флаги, карта доказательств.
- 7 Отчёт и решение о подписи.** HTML для врача + JSON для МИС, включая `send_gate`: разрешена ли ЭЦП при политике учреждения.

4.3. Восемь блоков итоговой оценки (веса)

Итоговый процент - **взвешенное среднее** по заполненным блокам. Если по блоку нет данных - его вес перераспределяется (не штрафует за «неизвестно»).

Блок	Вес	Что проверяется	Пример замечания
Оформление КЗ	15%	Дата, ФИО врача, специальность, заполненность разделов	«Нет даты консультации»
Данные пациента	10%	ФИО, возраст/ДР, пол	«Не указан возраст»
Применимость протокола	15%	Найден ли подходящий КП по МКБ и возрасту	«Протокол не подобран»
Диагноз	20%	Соответствие диагноза протоколу; подозрительный диагноз без дообследования	«Подозрение без маршрутизации»
Обследования	15%	Обязательные и рекомендованные исследования из протокола	«Не назначено УЗИ по протоколу»

Блок	Вес	Что проверяется	Пример замечания
Лечение	15%	Препараты, дозы, длительность vs протокол	«Доза не совпадает с КП»
Безопасность	5%	Красные флаги: онкология, тромбоз, тяжёлая инфекция...	«Признак без плана действий»
Контроль	5%	Сроки повторного визита, динамическое наблюдение	«Нет срока контроля»

4.4. Статусы итога

Статус	Обычно при балле	Смысл для врача
соответствует	≥ 90%	КЗ в целом согласовано с протоколом
в основном соответствует	75-89%	Мелкие пробелы
частично соответствует	50-74%	Нужна доработка перед отправкой
не соответствует	< 50%	Серьёзные расхождения с протоколом
нужен ручной разбор	любой	Критический red flag без ответа в КЗ
низкая уверенность	любой	Плохо распознан текст (скан, обрывки)

4.5. Карта доказательств (evidence_map)

Для методслужбы: таблица «правило протокола → найдено / не найдено в КЗ → цитата из PDF протокола». Это основа прозрачности - не «модель так решила», а **конкретный пункт КП**.

4.6. Ограничение балла по безопасности (safety cap)

Если в тексте найден опасный признак (тромбоз, онкология, тяжёлая инфекция...) без плана действий в рекомендациях - **верхняя цифра «Ориентировочное соответствие»** принудительно снижается (до 35-50% в зависимости от тяжести). Структурный отчёт (79% в примере ниже) при этом остаётся отдельной строкой - врач видит оба числа.

4.7. Уверенность разбора (confidence)

Отдельный показатель: насколько система уверена, что правильно поняла PDF. Низкая уверенность - повод не блокировать автоматически, а отправить на ручную проверку.

4.8. Конвейер L2: две ветки

4.8.1. L2-lite (production Render, default)

```
PDF / Bundle / текст
├─ L1 structured + alignment (CONSULT_ALIGNMENT_ENABLED=1)
├─ preload чанков только по matched paths (consult_memory, forbid full corpus)
├─ consult_evidence_pack + consult_evidence_quality (фильтр шума PDF)
├─ build_l2_fast_review (CONSULT_L2_FAST=1) ИЛИ LLM synthesize
├─ опц. CONSULT_L2_NARRATIVE / consult_criteria_narrative
├─ cisz_readiness + gate_score → send_gate
└─ UI: alignment + «Сверка с протоколом» (evidence pack)
```

4.8.2. L2 full (локально, CONSULT_RENDER_L2_LITE=0)

```
PDF / Bundle / текст + опционально рубрики каталога
├─ 1. Извлечение: PDF → text_extract; Bundle → fhir_bundle_adapter
├─ 2. Парсинг → ConsultationDocument; МКБ из блока «Диагноз»
├─ 3. match_protocol_cards + respiratory_path_filters (ОРВИ)
├─ 4. RAG retrieve: vector → lex+BM25 → embed-rerank (allowlist)
├─ 5. rule_checker + rules_summary_fallback → rules_compliance_pct
├─ 6. alignment + опц. LLM synthesise
├─ 7. analyze_consultation_text → 8 блоков → evidence_map, safety cap
├─ 8. Гибрид + cisz_readiness → gate_score → send_gate
└─ 9. JSON + UI
```

4.9. Три процента на экране - не путать (реализовано в UI)

На экране врача - **три карточки** с пастельной подсветкой по уровню (≥75 / 50-74 / <50). Длинные тексты - в спойлерах (<details>).

Карточка UI	Пример pl_1_f	Пример pl_2_d_s (СКВ)	Откуда считается	Для send_gate
Итог для врача (гибрид)	50%	58%	75%·структурный + 25%·правила; затем safety cap	Участвует в gate_score
Структурный разбор	79%	77%	8 взвешенных блоков (оформление, диагноз, лечение...)	Участвует в gate_score
Среднее LLM	40%	~38%	Среднее по критериям модели (пояснения, цитаты)	Нет - только для методслужбы
gate_score (в JSON)	50%	58%	min(гибрид, структурный) - compliance_gate.resolve_gate_score	Да - единый порог для МИС

Для «Айболит»: при pl_1_f врач видит гибрид 50%, структурный 79%, но gate_score = 50 - подпись при hard_gate с порогом 70% блокируется. При pl_2_d_s гибрид и структурный расходятся (58% vs 77%) - gate берёт минимум (58%).

4.10. Разбор реального примера: pl_1_f.pdf (флеботромбоз, I80.1)

Клиника: мужчина 48 лет, флеболог, флеботромбоз правой НК в стадии реканализации, ривароксабан, компрессия 2 класса, контроль УЗИ через 3 месяца.

Что сработало хорошо:

- Оформление и данные пациента - 100%.
- RAG подобрал верные PDF: КП по ТГВ (№17) и хроническим заболеваниям вен (№55).
- LLM: компрессия 100% с цитатой из протокола; жалобы 80%; объективный статус 60% (таблица Wells).
- Структурно: диагноз 80%, лечение 53%, контроль 55% - детализация по назначениям.
- Уверенность разбора 89% - текст PDF читается нормально.

Ограничения и артефакты (нужно знать методслужбе):

- **Онко-баннер** - ложное срабатывание: слово «рак» встречается в *тексте протокола по тромбозу* (риск при онкологии), а не в заключении. Это эвристика по фрагментам RAG, не клинический диагноз.
- **«Подобренные протоколы»** (карточки каталога) - попали КП по сердечной недостаточности: матчер карточек по рубрике I80 шире, чем RAG-отбор. Врачу ориентир - блок «Протоколы по тексту заключения» (PDF со страницами).
- **Диагноз 0% у LLM vs 80% структурно** - разные движки: модель учла жёсткие правила структуры диагноза из чужого КП (сердечная недостаточность: «нозология, функциональный класс»); в КЗ флеботромбоза локализация и стадия реканализации уже есть.

- **Фармакотерапия и наблюдение 0% у LLM** - в переданных *выдержках* RAG нет доз антикоагулянтов и графика УЗИ; это ограничение объёма контекста, не обязательно ошибка врача.
- **Правила протокола 0%** - rule_checker не нашёл срабатываний по каталогу; при наличии МКБ может подтянуться summary-fallback (до 1-2 карточек). Иначе гибрид тянет headline вниз.
- **Красные флаги** - «флеботромбоз» и «ривароксабан» помечены как частично учтённые (ожидаемо для данного диагноза).
- **Профиль рубрики 8,3%** - низкое совпадение с карточками «болезни системы кровообращения» из-за неверных карточек; не путать с клиническим качеством КЗ.

4.11. Блоки интерфейса (компактный дашборд, июнь 2026)

Блок UI	Видимость	Содержание
Дашборд оценок	Сразу	3 карточки: гибрид, структурный, LLM (если есть); донаты и таблицы блоков
Карточки согласования	Сразу (L1/L2)	Alignment по блокам: КЗ, КП, СОП Кравира, gaps, criteria_source
Сверка с протоколом	L2-lite	Evidence pack: выдержки КП с читаемыми названиями, фильтр оглавления PDF
Решение о подписи	Сразу	sign_decision / sign_decision_ru - одна рекомендация врачу (подписать / доработать / override)
Готовность к ЦИСЗ	Сразу (отдельный блок)	Чек-лист «Готовность к ЦИСЗ - Первичный приём (3.2.1)»: ✓/X по проверкам, критические пробелы, спойлер «Полная расшифровка» (cisz_readiness, слои A-C)
Краткий вывод	Спойлер (открыт)	summary_ru от LLM
Ограничения оценки	Спойлер	limitations_ru (нет КП в RAG, неполные выдержки...)
Допуск к подписи (МИС)	Сразу	send_gate: статус, gate_score, порог 70%
Сводка методслужбы	Спойлер	Зоны риска <70% с чипами %; группы 70-84% / ≥85%
Структурный отчёт	Спойлер	HTML: A/B/C, 8 блоков, диагноз/лечение, red flags, evidence_map - пастельные бейджи
Критерии модели	Спойлер	LLM-критерии (L2 full) или дубль alignment; цитаты - во вложенных «Цитаты»
Протоколы · N PDF	Спойлер	Чипы PDF; обрезанные фрагменты разделов
Онко-предупреждение	Сразу (если есть)	Только по тексту КЗ, не по фрагментам протоколов
Экспорт	Кнопки	Печать, HTML методслужбы, JSON (send_gate)

4.12. Разбор примера: pl_2_d_s.pdf (дискоидная красная волчанка L93.0)

Клиника: мужчина 48 лет, дерматолог, предварительный диагноз «Дискоидная красная волчанка ?», гидроксихлорохин, Тридерм, УФ-защита, контроль через 2 недели.

- **Итог 58%** - гибрид с учётом rules 0% (нет срабатываний по фрагментам RAG для L93.0).
- **Структурный 77%** - оформление 100%, диагноз 70% (предварительный «?», неполная структура формулировки).

- **LLM ~38%** - низко по лечению/наблюдению: в выдержках RAG нет специфического КП по L93.0, подобраны общие дерматологические PDF.
- **gate_score 58%** - ниже порога 70%; в режиме inform - предупреждение, в hard_gate - блокировка.
- **Урок для корпуса:** нужен КП по СКВ/дерматовенерологии с L93.0 в индексе; иначе клинически верное лечение получает низкий % из-за отсутствия цитат в RAG.

5. FHIR BY, ЦИСЗ и место Protocol в контуре МИС

Protocol встраивается между заполнением КЗ врачом и подписью ЭЦП. Он не заменяет МИС и не вызывает ЦИСЗ - даёт врачу и методслужбе понятный ответ: «можно ли подписывать» и «что исправить».

Нормативная база: стандарт [FHIR BY](#) (FHIR 5.0), Протокол взаимодействия МИС ОЗ с ЦИСЗ v.1.4, программа испытаний МИС v.1.3-4 (амбулаторный профиль). Подробные спецификации - на [fhir.by](#).

5.1. Зачем Protocol, если есть ЦИСЗ?

ЦИСЗ отвечает на вопрос: «Правильно ли упакованы данные?» - структура пакета, профили FHIR, коды из справочников НСИ. Это техническая валидация **после** подписи и импорта.

Protocol отвечает на другие вопросы **до** подписи:

- **Клиника:** соответствует ли содержание заключения клиническим протоколам Минздрава РБ? (диагноз, обследования, лечение, red flags - с цитатами из официальных PDF).
- **Готовность пакета:** собрал ли МИС все обязательные поля, которые ЦИСЗ отклонит при импорте? (пациент, визит, диагноз, автор, организация).
- **Единый порог:** $gate_score = \min(\text{клиника}, \text{готовность ЦИСЗ})$ - одна цифра для политики подписи в МИС.

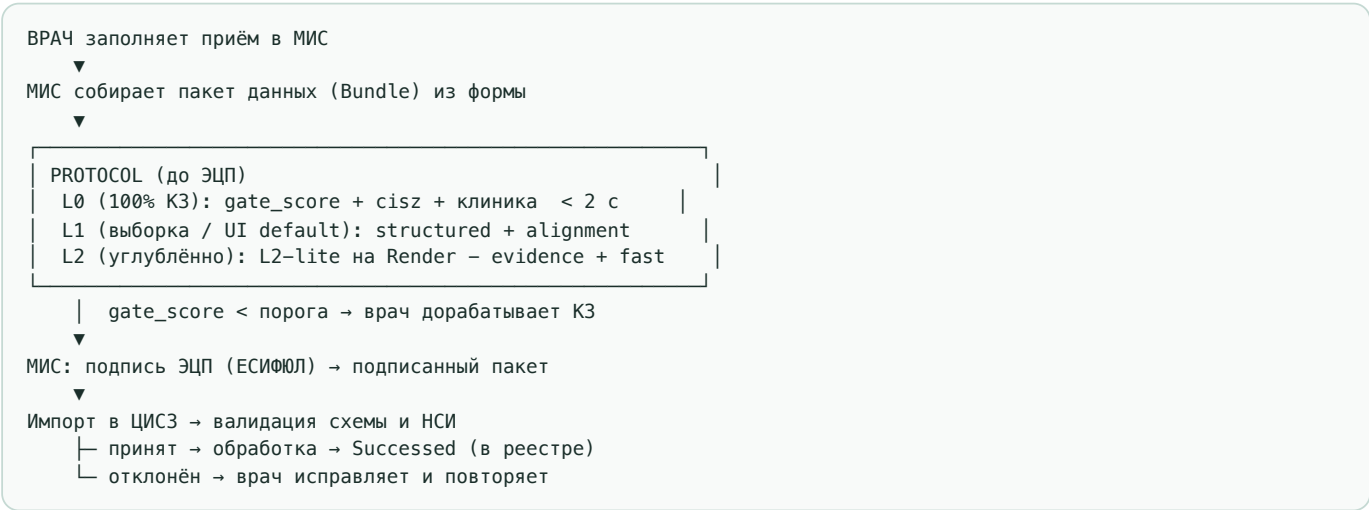
Польза для медцентра: меньше отклонённых пакетов в ЦИСЗ, меньше слабых КЗ в государственном реестре, прозрачная обратная связь врачу до нажатия «Подписать». Protocol не является МЭЭ и не блокирует юридически - политика настраивается в МИС.

5.2. Три системы - кто за что отвечает

Система	Роль	Что проверяет	Когда
МИС «Айболит»	Рабочее место врача	Форма приёма, сбор пакета данных, ЭЦП, вызов Protocol	При сохранении КЗ
Protocol	Клинический аудитор + предпроверка пакета	КП Минздрава + готовность Bundle к ЦИСЗ	До ЭЦП
ЦИСЗ	Государственный контур	Профили FHIR, НСИ, подпись, импорт	После ЭЦП

Пакет данных (Bundle) - это «конверт»: внутри пациент, визит, диагноз, осмотр, назначения. МИС собирает конверт; ЦИСЗ проверяет, что конверт оформлен по правилам; Protocol проверяет, что *содержание* заключения согласовано с протоколами лечения Минздрава.

5.3. Жизненный цикл: от приёма до реестра Минздрава



5.4. Два контура Protocol - клиника и ЦИСЗ

Контур	Вопрос	Модуль	ЦИСЗ это проверяет?
Клинические протоколы	Соответствует ли лечение и диагноз КП Минздрава?	compliance_engine, rule_checker, RAG	Нет
Готовность к ЦИСЗ	Не отклонят ли пакет при импорте?	cisz_readiness	Да (после подписи)

Итог для МИС: `gate_score = min(клинический балл, cisz_score)`. Врач видит оба числа отдельно.

5.5. Что внутри пакета данных (простыми словами)

Блок данных	Смысл для врача	Зачем Protocol
Пациент	Кто лечится (ИНП, пол, возраст)	Блок «Данные пациента», применимость КП
Визит	Когда и где был приём	Оформление, привязка диагноза к визиту
Диагноз (МКБ-10)	Код и формулировка	Подбор протоколов (40% веса матчера)
Осмотр, vitals	Жалобы, статус, АД, температура	8 блоков, red flags
Назначения	Препараты, обследования	Блоки «Лечение», «Обследования» vs КП
Автор и организация	Кто подписывает и от чьего имени	cisz: обязательны для импорта

Адаптер `fhir_bundle_adapter.py` переводит Bundle в ConsultationDocument - тот же путь анализа, что и для PDF.

5.6. Три слоя предпроверки ЦИСЗ (cisz_readiness)

Слой	Что проверяем	Типичная ошибка
А. Обёртка пакета	Тип document, профиль пакета, идентификатор, дата	Неверный тип → отклонение при импорте
В. Документ (Composition)	Первым в пакете; автор, организация, пациент, визит	Нет ссылки на врача → ошибка импорта
С. Клинические ресурсы	Пациент, визит, диагноз, vitals, назначения (сценарий 3.2.1)	Пустой диагноз → слабое КЗ в реестре

Protocol проверяет *наличие* обязательных полей локально за < 2 с. Серверная валидация НСИ и ЭЦП выполняется ЦИСЗ после подписи. Сверку author/custodian с OAuth-токеном делает МИС перед импортом.

5.7. API Protocol для МИС

Endpoint	Уровень	Когда вызывать	Ключевые поля ответа
POST /api/consult-compliance-screen	L0	На каждое сохранение КЗ	send_gate, gate_score, cisz_readiness, <2 с
POST /api/consult-review/tier	L0/L1/L2	Явный выбор глубины проверки	Как выше или полный отчёт (L1/L2)
POST /api/consult-validate-bundle	ЦИСЗ only	Отладка Bundle без клиники	cisz_readiness
POST /api/consult-review/json	L2	Углублённый разбор, выборка	review, structured_analysis, RAG-цитаты

Endpoint	Уровень	Когда вызывать	Ключевые поля ответа
POST /api/consult-review	L2 (PDF)	Пилот / UI	Тот же JSON + HTML

5.8. Protocol vs ЦИСЗ - сравнение для руководства

Критерий	Protocol (до ЭЦП)	ЦИСЗ (после ЭЦП)
Соответствие КП Минздрава	Да - 8 блоков, evidence_map, цитаты PDF	Нет
Готовность пакета к импорту	Да - cisz_readiness за < 2 с	Да - полная валидация схемы и НСИ
Клиническое качество текста КЗ	Да - диагноз, лечение, red flags	Нет
Электронная подпись	Нет	Обязательно
Время ответа врачу	Секунды (L0)	Десятки секунд + опрос статуса

Итог для интеграции: МИС вызывает Protocol при сохранении → получает gate_score и список замечаний → врач исправляет → только затем ЭЦП и импорт в ЦИСЗ. Так снижается доля отклонённых пакетов и слабых КЗ в реестре Минздрава.

5.17. Чек-лист перед подписью ЭЦП (для врача)

Одностраничная памятка для рабочего места врача и методслужбы: [pre-sign-checklist-print.html](#) (печатать / PDF).
Ниже - та же логика внутри архитектурного документа.

Сохранить в МИС → Protocol L0 (send_gate) → gate OK? → ЕСИФЮЛ → CMS → ЦИСЗ
gate_score = min(клиника по КП, cisz_readiness) · порог учреждения обычно 70%

Три уровня проверки перед подписью

Уровень	Что проверить	Пример ошибки
Критично	Пациент (INP), МКБ-10 = текст диагноза, не «?» как заключительный, author/custodian = вы и ваше ОЗ, диагноз привязан к визиту	Чужой пациент; импорт отклонён; ложный final в ЦИСЗ
Важно	Жалобы, осмотр, назначения (доза, курс), контроль, red flags с маршрутом, направление при консультации	Слабое КЗ в реестре; претензии методслужбы / по КП
Формально	Encounter completed, vitals/anthropo, Bundle.type=document, Composition первым в entry	Низкий cisz_score; ProcessingStatus Failed

Соответствие: поле FHIR Bundle → форма врача → Protocol

Поле FHIR	В форме МИС	Риск	Контуп Protocol
Patient.identifier (INP)	Карта пациента	крит	cisz · Patient
Condition.code (МКБ-10)	Код и формулировка диагноза	крит	8 блоков · диагноз + КП
KindOfDiagnosis / final	Предварительный vs заключительный	крит	FinalDiagnosis
Composition.author	Подписывающий врач	крит	cisz слой B (= OAuth-токен)
Composition.custodian	Медорганизация ОЗ	крит	cisz слой B (= OAuth-токен)
Encounter.diagnosis	Диагноз визита	крит	cisz 3.2.1
ObservationSubjective	Жалобы	высок	cisz + оформление
ObservationObjective	Объективный статус	высок	8 блоков · структура
VitalSignsBy / anthropo	АД, ЧСС, рост, вес	средн	cisz 3.2.1
MedicationRequest	Назначения	высок	8 блоков · лечение + КП
Контроль / рекомендации	Срок повторного визита	высок	8 блоков · контроль
ServiceRequest	Направление (консультация)	высок	cisz 3.13.1
Bundle.type, Composition first	Формирует МИС	средн	cisz слои A-B

Два контура: ЦИСЗ проверяет схему и НСИ; клинические протоколы лечения на импорте не оценивает. Protocol закрывает оба риска до ЭЦП. Памятка на одну страницу: [pre-sign-checklist-](#)

6. Блокировка отправки при низкой оценке (send_gate)

Риск для медцентра: врач подписал слабое КЗ → данные в ЦИСЗ → при проверке методслужбой Минздрава видны пробелы. **Protocol не отправляет данные в ЦИСЗ** - он возвращает в «Айболит» рекомендацию, можно ли нажимать «Подписать».

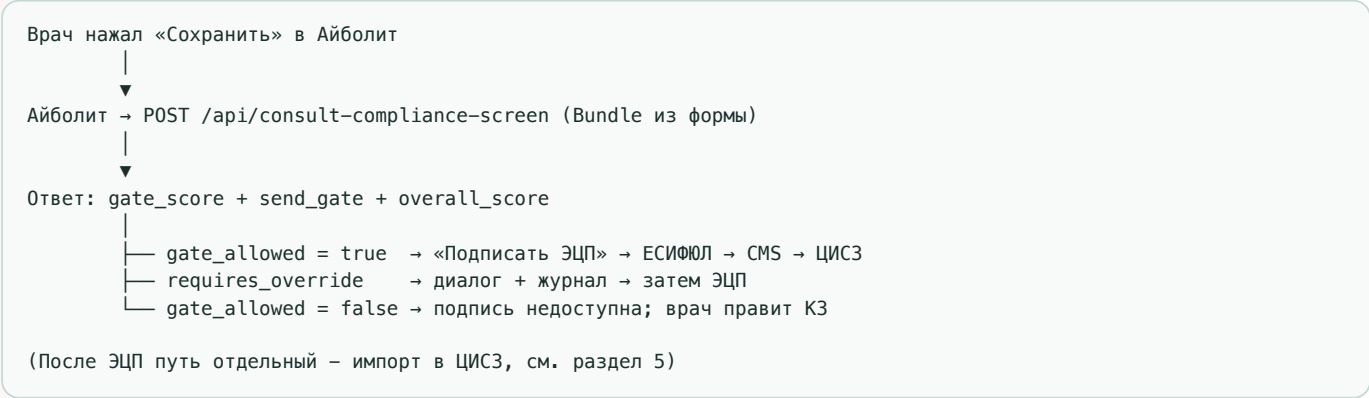
В отчёт и API добавлен блок send_gate (compliance_gate.py):

Поле	Смысл
gate_score	Единый балл для порога: min(клинический, cisz_score), где клинический = min(гибрид, структурный)
cisz_score / cisz_readiness	Готовность FHIR-пакета к импорту (чек-лист 3.2.1 / 3.13.1 / 3.3)
headline_score / structural_score	Компоненты для UI и аудита
gate_allowed	false - «Айболит» не вызывает ЭЦП (при hard_gate)
requires_override	true - только «Подписать всё равно» + журнал
send_risk_level	low / medium / high / blocked
block_reason_ru	Текст для врача (порог 70%, критические замечания)
disclaimer_ru	Не МЭЭ, не юридическое заключение

Переменные окружения: COMPLIANCE_GATE_MODE (по умолчанию **soft_gate**; также inform | hard_gate | critical_only), COMPLIANCE_GATE_MIN_SCORE (70), COMPLIANCE_GATE_SOFT_MIN_SCORE (55). Критические пробелы cisz_readiness в soft_gate требуют подтверждения врача; в hard_gate блокируют подпись.

6.1. Режимы политики (настраивает учреждение в «Айболит»)

Режим	Поведение	Когда применять
inform	Только предупреждение, подпись всегда доступна	Пилот, обучение врачей
soft_gate	При балле < порога (по умолчанию 70%) - диалог «Подписать всё равно?»	Рабочий режим поликлиники
hard_gate	Подпись заблокирована при низком балле или критических замечаниях	Ужесточение после пилота
critical_only	Блок только при red flags безопасности	Минимальное вмешательство



Формулировка для врача всегда: *ориентир методслужбы по клиническим протоколам*, не МЭЭ и не автоматическая санкция.

7. Масштаб: тысячи КЗ в сутки

Полная проверка с LLM на каждое заключение не масштабируется. Решение - **этажи**:

Уровень	Охват	Содержание	Время
L0	100% КЗ	/api/consult-compliance-screen или tier=L0: parse + 8 блоков + cisz + send_gate	< 2 сек
L1	10-20% / UI default	Structured + alignment + СОП + rich chunks; без full corpus RAG	~5-15 с
L2	2-5% / кнопка L2	Render: L2-lite (evidence pack + L2-fast). Local: full RAG + LLM	~15-30 с / 30-90 с
L3	0,5-1%	Ручной разбор методслужбой	человек

«Айболит» на каждое сохранение вызывает **L0** (мгновенный send_gate). L2 - по кнопке методиста или автоматически для «красной зоны».

8. Интеграция с «Айболит» - план по этапам

Этап	Статус	Protocol	Айболит (EPAM)
A	Готово (backend)	send_gate, gate_score, JSON-отчёт, UI-пилот PDF	Кнопка «Проверить»; встраивание панели gate
B	Готово (backend)	FHIR-адаптер, L0/L2 API, cisz_readiness, validate-bundle	Автовызов L0 при сохранении; soft_gate перед ЕСИФЮЛ
C	В работе	Composition v1.4 в cisz_readiness; прокси \$validate (опц.)	Журнал override; ProcessingStatus после импорта
D	План	Аналитика методслужбы (агрегаты без ПДн)	Дашборд качества; связка gate ↔ ProcessingStatus Failed

Справочно: программы испытаний МИС v.1.3 (части 1-6: скорая, МСЭ, патология, амбулаторный, стационар, стоматология) - на fhir.by/interactionMIS; для «Айболит» релевантна **часть 4 - амбулаторный профиль**.

9. Состояние проекта (июнь 2026)

Реализовано (июнь 2026)

- `compliance_gate.py` - `send_gate`, `gate_score`, режимы `inform/soft/hard/critical_only`, `sign_decision`.
- API: `L0 /api/consult-compliance-screen`, `L1/L2 /api/consult-review/tier`, `stream /api/consult-review/stream`.
- **Alignment:** `consult_alignment.py`, `consult_criteria_enrichment.py`, `kravira_sop_rules.py` - primary criteria в L1/L2-lite.
- **L2-lite на Render:** `consult_evidence_pack.py`, `consult_l2_review.py`, `consult_memory.py`, `CONSULT_RENDER_L2_LITE=1`.
- **Lazy RAG:** `manifest` + `disk` (`RAG_LAZY_CHUNK_STORE`), `path` `lex shards`, `forbid full corpus retrieve`.
- Семантический RAG: `embeddings`, `vector index`, `semantic rules`, `tiering`, `batch enrichment`.
- `fhir_bundle_adapter.py`, `cisz_readiness.py`, `fhir_mis_test_matrix.py` - FHIR BY + программа испытаний v.1.3-4.
- `respiratory_path_filters.py` - гигиена подбора КП при ОРВИ (J00-J06).
- `plan_definition_export.py` - черновик экспорта `Protocol Summary` → FHIR `PlanDefinition` (CDS hook).
- Методслужба: `feedback_store.py`, `Methodist workbench` в UI, `/api/ml/feedback`.
- Поиск: `tiers S0/S1/S2`, воронка `search_funnel.py`, чек-лист оформления КЗ по СОП в карточке рекомендованного КП.
- Гибридный headline: 75% структурный + 25% правила + `safety cap`; онко-скан только по тексту КЗ.
- UI: `alignment cards`, `evidence pack`, ЦИСЗ 3.2.1, `send_gate`, `Methodist panels`.
- Тесты: `alignment`, `criteria enrichment`, `evidence quality`, `cisz`, `compliance_gate`, `respiratory filters`.

В работе / следующий приоритет

1. **Интеграция в «Айболит»** - вызов L0 при сохранении формы; блокировка ЕСИФЮЛ по `gate_allowed`.
2. **Сценарии ЦИСЗ** - стационар/стоматология (программа испытаний ч.5-6); амбулаторный 3.2.1 уже в UI-чек-листе.
3. **Сверка `author/custodian` с OAuth-токеном** - в МИС перед импортом (`Protocol` проверяет только наличие ссылок).
4. **Корпус L93.0 / дерматовенерология** - специфические КП в RAG (пример `pl_2_d_s`: `rules 0%` из-за отсутствия выдержек).
5. **Правила структуры диагноза** - профили для ТГВ (локализация, стадия) и дерматологии (форма, локализация без «?»).
6. **Журнал `override`** в МИС - кто подписал при `gate_score` < порога.
7. **Связка с ЦИСЗ** - после импорта показывать `ProcessingStatus` рядом с результатом `Protocol` (разные оси качества).
8. **Валидация `Bundle` без импорта** - опционально проксировать метод ЦИСЗ для отладки до ЭЦП (на стороне МИС).

Ограничения (явно)

- `Protocol` **не подписывает** `Bundle` и **не импортирует** в ЦИСЗ.
- Оценка по протоколам Минздрава - ориентир методслужбы, не замена МЭЭ.
- При отсутствии КП в корпусе низкий % LLM/rules не всегда означает ошибку врача (см. `pl_2_d_s`).

10. Карта модулей (для технического аудита)

Модуль	Назначение	Вход	Выход
rag_server.py	HTTP API, версия BUILD_VERSION	REST / multipart	JSON, SSE stream
consult_review_pipeline.py	L2-lite / full L2 с SSE-прогрессом	text / bundle meta	review + alignment + evidence + cisz
consult_alignment.py	Карточки согласования КЗ ↔ КП ↔ СОП	ConsultationDocument + paths	alignment[], criteria
consult_criteria_enrichment.py	Выдержки и покрытие по блокам	K3 + chunks + summary	enriched cards
kravira_sop_rules.py	Эталон СОП №2 (doc_kravira)	block_id	findings/gaps, reference
consult_evidence_pack.py	Выдержки КП для L2-lite UI	matched paths	evidence_map, protocol_ui_meta
consult_l2_review.py	L2-fast deterministic review	L1 + evidence	block_gaps, l2_mode
consult_memory.py	Кэш чанков, preload paths	PDF paths	chunk cache, forbid full corpus
respiratory_path_filters.py	Фильтр ложных КП при ОРВИ	ICD J00-J06	filtered paths
plan_definition_export.py	Summary → FHIR PlanDefinition	protocol summary card	draft CDS resource
feedback_store.py	ML-разметка методслужбы	JSONL events	export tar.gz
rules_summary_fallback.py	Доп. правила из summary по МКБ	rules_check, ICD	обновлённый rules_compliance_pct
protocol_summary/icd_index.py	Быстрый индекс МКБ → summary	ICD-10	refs для fallback; prewarm при старте
consult_screen.py	L0 скрининг (<2 с)	text bundle	send_gate + cisz_readiness
consult_tiering.py	Маршрутизация L0/L1/L2	tier + text/bundle	screen structured delegate L2
protocol_match.py	Подбор карточек КП	МКБ, диагноз, рубрика	match_score, applicability
consult_overall_score.py	Гибрид headline %	structured + rules	75/25 blend, safety cap
vector_index.py	FAISS/numpy prefilter RAG	query embedding	top-K chunk indices

Модуль	Назначение	Вход	Выход
semantic_rule_fallback.py	Синонимы в правилах	required_exam, keyword	alias/fuzzy match
compliance_engine.py	8 блоков vs КП	ConsultationDocument	ComplianceReport
compliance_gate.py	Политика подписи	ComplianceReport	send_gate, gate_score
cisz_readiness.py	Готовность к ЦИСЗ	Bundle text	%, checks[], merge gate
fhir_bundle_adapter.py	FHIR → ConsultationDocument	Bundle	текст + структура
fhir_bundle_inspect.py	Разбор Bundle	Bundle	bool-мапа проверок
fhir_mis_test_matrix.py	Весы и критичность чеков	scenario id	MisCheckDef[]
consult_report.py	HTML отчёт методслужбы	ComplianceReport	HTML + JSON

10.1. Автотесты (pytest)

- tests/test_cisz_readiness.py - слои v1.4 + сценарий 3.2.1, merge gate, расшифровка
- tests/test_icd_summary_index.py, tests/test_rules_summary_fallback.py - быстрый fallback
- tests/test_compliance_gate.py - режимы inform/soft/hard/critical_only
- tests/test_fhir_bundle_adapter.py - маппинг Bundle → K3
- tests/test_consult_screen.py, test_consult_tiering.py - L0/L1/L2
- tests/test_consult_alignment.py, test_consult_criteria_enrichment.py, test_consult_evidence_quality.py
- tests/test_respiratory_path_filters.py, test_plan_definition_export.py
- Фикстуры FHIR: tests/fixtures/fhir_by/

10.3. B2C-контур для пациентов (production)

Отдельный клиентский shell patient.html («Проверь КЗ») использует тот же детерминированный L1-pipeline, но отдаёт пациенту sanitize-отчёт без полей B2B (gate_score, cisz, methodist). White-label клиники, tier-цены, PWA + service worker.

- Frontend: patient.html, patient-ui.js, sticky mini logo; B2B index.html - mini-logo + pastel pill tabs в одной строке.
- API: POST /api/patient/review, /api/patient/payment/session, GET /api/patient/config.
- Документация: docs/architecture-b2c-patient.md, настройки monetization в кабинете методиста (#b2c-monetization).

10.4. Нормативные документы в репозитории

- data/Программа испытаний МИС v.1.3-4 – Амбулаторный профиль.pdf
- data/Протокол информационного взаимодействия МИС 03 с ЦИСЗ_v_1.4.pdf
- Публично: fhir.by/interactionMIS, [профили FHIR BY](#)

11. Заключение для методслужбы и ИТ Минздрава

Protocol реализует **трёхконтурную предимпортную экспертизу** амбулаторного заключения:

1. **Структурный контур** - 8 блоков compliance, evidence_map, safety cap, send_gate.
2. **Alignment** - детерминированное согласование КЗ с клиническими протоколами Минздрава и СОП №2 Кравира (оформление амбулаторной карты).
3. **Контур готовности к ЦИСЗ** - зеркало требований Протокола взаимодействия v.1.4 и программы испытаний МИС v.1.3-4.

Итоговый gate_score = min(клинический балл, cisz_score) даёт МИС единый порог для политики подписи.

Система **прозрачна**: каждая оценка раскладывается на проверки с весами; критические пробелы по ЦИСЗ и red flags по безопасности учитываются отдельно.

Protocol **не подменяет** ЦИСЗ: серверная валидация НСИ, OperationOutcome и ProcessingStatus остаются на стороне государственного контура. Задача модуля - снизить долю отклонённых пакетов и клинически слабых КЗ до вызова ЕСИФЮЛ и Bundle/\$import.

МЦ «Кравира» · Protocol · МИС «Айболит» (ЕРАМ) · [FHIR BY](#) · Протокол МИС ОЗ-ЦИСЗ v.1.4 · Программа испытаний МИС v.1.3-4 (амбулаторный) · КП Минздрава РБ