Техническое задание: модуль медицинской безопасности ИИ-ассистента для беременных и молодых мам (казахский, русский, английский)

Nov 15, 2026

1. Общие сведения

Нужен модуль, который до показа пользователю определяет срочность медицинского обращения и проверяет каждое медицинское утверждение в ответе ИИ по проверенной базе знаний — на казахском, русском и английском.

Параметр	Значение
Заказчик	Частная компания «Supermoms Ltd.», г. Астана
Продукт заказчика	Supermoms Club — мобильное приложение для беременных и молодых мам (iOS, Android): календарь беременности по неделям, сообщество вопросов и ответов, дневники, ИИ-ассистент; интерфейс на нескольких языках, в том числе казахском, русском и английском
Наименование работ	Модуль медицинской безопасности ИИ-ассистента: триаж срочности обращений и проверка медицинских утверждений в ответах ИИ
Вид работ	НИОКР с созданием программного обеспечения: модели, сервис с API, база знаний, тест-набор
Срок	до 22 недель, три этапа (раздел 9)
Где работает результат	на серверах заказчика; обращения пользователей не передаются третьим лицам
Вознаграждение	по договорённости после оценки технического и коммерческого предложения

Модуль не ставит диагнозов и не назначает лечение. Его задача — не пропустить экстренную ситуацию и не допустить к показу непроверенный или опасный медицинский совет.

2. Цель работ

Цель — каждое обращение с признаками опасного состояния при беременности, после родов или у ребёнка до 3 лет должно получать первой фразой правильную маршрутизацию (скорая, приёмный покой, врач сегодня), а ни одно медицинское утверждение ИИ не должно доходить до пользователя без подтверждения источником.

Результат состоит из пяти частей:

1. **Триаж срочности.** Классификатор текста обращения по четырём уровням: экстренно / срочно к врачу / планово / не медицинский вопрос, с перечнем найденных «красных флагов». Работает на казахском, русском, английском и их смеси, с разговорной лексикой, опечатками и латиницей.
2. **Проверка утверждений.** Сервис выделяет из черновика ответа ИИ медицинские утверждения, сверяет каждое с базой знаний и возвращает вердикт: подтверждено источником / не подтверждено / противоречит источнику / запрещённый тип (дозировка, диагноз, отмена препарата).
3. **База знаний.** Структурированные положения клинических протоколов МЗ РК по ведению беременности, послеродовому периоду и наблюдению детей раннего возраста и рекомендаций ВОЗ; у каждого положения — источник, дата редакции и текст на трёх языках.
4. **Сервис с REST API** для развёртывания на серверах заказчика, без передачи текстов обращений внешним сервисам; не привязан к конкретной модели, которая пишет ответы.
5. **Закрытый тест-набор и методика оценки**, размеченные практикующими врачами, — чтобы качество можно было измерить при приёмке и перепроверять при каждой смене модели.

3. Проблема и обоснование

Универсальные языковые модели отвечают на медицинские вопросы уверенно, но не гарантируют ни распознавания экстренной ситуации, ни соответствия совета клиническим протоколам. В акушерстве и педиатрии цена такой ошибки — потерянные часы при кровотечении, преэклампсии или лихорадке у новорождённого.

Почему задача не решается списком ключевых слов или инструкцией в промпте:

- **Язык.** Мамы пишут на смеси казахского и русского, латиницей и кириллицей, с опечатками и бытовыми словами вместо терминов. Казахский — малоресурсный язык для медицинской обработки текста: открытых размеченных наборов по акушерским и педиатрическим жалобам на нём нет.
- **Контекст меняет смысл.** «Кровотечения нет», «было в первом триместре, сейчас всё хорошо», «у подруги», «а что делать, если вдруг…» — те же слова без экстренности. И наоборот: «чуть-чуть мажет, но живот каменный уже час» — экстренность без ключевых слов.
- **Срок и возраст решают.** Температура 38 °C у ребёнка до 3 месяцев и в 2 года — разная срочность; слабые шевеления до 20 недель и после 28 — тоже. Срок беременности и возраст ребёнка есть в профиле приложения и должны учитываться.
- **Правдоподобная ошибка не видна.** Уверенный совет, противоречащий протоколу, читается как нормальный. Проверить его можно только сверкой с источником, а клинические протоколы МЗ РК не существуют в машиночитаемом виде и на казахском в форме, пригодной для сверки.

Типичные ошибки универсальных моделей на таких обращениях (примеры составлены для иллюстрации):

Обращение	Что делает модель	Риск
«30 апта, қан кетіп жатыр, что делать?»	Отвечает общими словами о выделениях, не поняв казахскую часть	Пропуск кровотечения в III триместре
«34 неделя, с утра болит голова, ноги опухли»	Связывает с усталостью, советует отдохнуть	Пропуск признаков преэклампсии
«Сыну 2 месяца, 38,4 и вялый»	Называет жаропонижающее и дозу	Задержка осмотра врачом; дозировка без врача
«Бала бүгін аз қимылдады, 36 недель»	Советует подождать до завтра	Пропуск снижения шевелений плода
«Можно ли при беременности препарат N?»	Уверенно называет его безопасным без источника	Вред плоду; самолечение

Готовых решений нет: общие сервисы модерации не оценивают медицинскую срочность, а медицинские модели и наборы данных сделаны в основном для английского и зарубежных протоколов. Передача жалоб о здоровье во внешние сервисы для проверки нежелательна: это особо чувствительные персональные данные, и их обработка должна соответствовать Закону РК «О персональных данных и их защите».

4. Функциональные требования

Модуль делает две проверки — обращения до ответа ИИ и ответа до показа — и опирается на базу знаний, которую ведут врачи.

4.1. Триаж срочности обращения

№	Требование
ФТ-1	Вход: текст до 4 000 символов + необязательный контекст: срок беременности в неделях или дата родов, возраст ребёнка в днях, страна, язык интерфейса
ФТ-2	Выход: уровень (экстренно / срочно к врачу / планово / не медицинский), уверенность, найденные красные флаги из справочника, фрагменты текста-основания, определённые языки
ФТ-3	Справочник красных флагов охватывает беременность, роды, послеродовый период, ребёнка 0–3 лет и психическое состояние матери, включая мысли о вреде себе или ребёнку; состав согласуется с врачами-экспертами на этапе 1
ФТ-4	Корректно обрабатываются отрицание, прошедшее время, третье лицо, гипотетические вопросы, смесь языков и алфавитов в одном сообщении
ФТ-5	При неуверенности между двумя уровнями выбирается более срочный; пороги настраиваются без переобучения

№	Требование
ФТ-6	При уровне «экстренно» сервис возвращает готовую первую фразу на языке пользователя: что сделать прямо сейчас и номер экстренной помощи страны из справочника заказчика
ФТ-7	Обрабатываются не только сообщения чата, но и посты и комментарии сообщества (длинный текст, несколько тем в одном сообщении)

4.2. Проверка ответа ИИ до показа

№	Требование
ФТ-8	Вход: обращение, черновик ответа ИИ, контекст и результат триажа. Выход: список медицинских утверждений с вердиктом и ссылкой на положение базы знаний + решение: показать / показать с правками / заменить безопасным шаблоном
ФТ-9	Запрещённые типы утверждений блокируются всегда: дозировка, начало или отмена препарата, диагноз, прогноз исхода, совет не обращаться за помощью
ФТ-10	Утверждение, противоречащее источнику, удаляется или заменяется формулировкой из базы знаний; неподтверждённое — удаляется или смягчается до «обсудите с врачом»; язык и тон ответа сохраняются
ФТ-11	Сверка с триажем: если обращение «экстренно», а ответ не начинается с указания, что делать прямо сейчас, ответ заменяется или дополняется первой фразой из ФТ-6
ФТ-12	Немедицинские части ответа (поддержка, бытовые советы) не блокируются и не переписываются

4.3. База знаний

№	Требование
ФТ-13	Положение = одно проверяемое утверждение + условия применимости (срок, возраст) + источник (документ, раздел, дата редакции) + формулировки на казахском, русском и английском + отметка врача, который его проверил
ФТ-14	Источники первой версии: действующие клинические протоколы МЗ РК по ведению беременности, послеродовому периоду и наблюдению детей раннего возраста; рекомендации ВОЗ по антенатальной и постнатальной помощи. Перечень документов утверждает заказчик
ФТ-15	Первая версия — не менее 800 положений, покрывающих все красные флаги справочника и темы тест-набора
ФТ-16	Веб-редактор для врача: добавить, изменить, отозвать положение; история изменений; после правки — автоматический прогон затронутых тестов
ФТ-17	База версионируется; каждый ответ сервиса содержит номер версии базы и моделей

4.4. Журнал и разбор спорных случаев

№	Требование
ФТ-18	Журнал решений: время, уровень, флаги, вердикты, версии. Текст обращения по умолчанию не хранится; хранение обезличенного текста включается настройкой, срок хранения настраивается
ФТ-19	Очередь разбора для врача: случаи с низкой уверенностью и отметки «неверно»; врач ставит правильный уровень, случай попадает в набор для дообучения

№	Требование
ФТ-20	Еженедельный отчёт: распределение уровней, доля заблокированных и исправленных ответов, наиболее частые неподтверждённые утверждения

5. Архитектура и интерфейс

Модуль стоит между приложением и моделью, которая пишет ответы, и вызывается дважды: до генерации ответа (триаж) и после неё, до показа (проверка). Модель ответов остаётся заменяемой: модуль не зависит от того, какая модель сделала черновик.

Ответ модели доходит до пользователя только после проверки

Пунктир — модуль, который разрабатывает исполнитель; цифры — порядок шагов



1 — обращение уходит на триаж; при «экстренно» приложение сразу показывает первую фразу
2 — модель ответов заказчика пишет черновик с учётом уровня срочности
3 — черновик сверяется с базой знаний; пользователь видит его, правленый текст или замену
4 — решения без текстов пишутся в журнал, спорные случаи уходят в очередь врача

Архитектура модуля · 6 компонентов модуля, 2 системы заказчика, 4 шага

Модуль не пишет ответы сам: он оценивает срочность обращения и проверяет готовый черновик, а при необходимости заменяет его безопасным текстом из базы знаний.

5.1. Программный интерфейс

REST API по HTTPS, JSON, описание в формате OpenAPI 3. Доступ по сервисному токену из внутренней сети заказчика.

Метод	Назначение
`POST /v1/triage`	уровень срочности обращения, красные флаги, первая фраза при экстренности
`POST /v1/verify`	проверка черновика ответа: утверждения, вердикты, решение, исправленный текст
`POST /v1/feedback`	отметка «неверно» от модератора или врача → очередь разбора
`GET /v1/health`, `GET /v1/version`	состояние сервиса, версии моделей и базы знаний

Пример запроса триажа (данные вымышленные):

```
{
  "text": "30 апта, қан кетіп жатыр, что делать?",
  "channel": "chat",
  "context": {"pregnancy_week": 30, "child_age_days": null, "country": "KZ",
"ui_lang": "ru"}
}
```

Ответ:

```
{
  "level": "emergency",
  "confidence": 0.97,
  "red_flags": [{"code": "PREG_BLEEDING_T3", "span": "қан кетіп жатыр"}],
  "languages": ["kk", "ru"],
  "first_line": "Кровотечение на 30-й неделе — повод для срочной помощи:
прямо сейчас вызовите скорую по номеру 103.",
  "kb_version": "1.0.0",
  "model_version": "triage-1.0.0"
}
```

Ответ проверки (`/v1/verify`) содержит `decision` (`show` / `show_edited` / `replace`), список `claims` — у каждого текст, `verdict` (`supported` / `unsupported` / `contradicts` / `forbidden`) и ссылка на положение базы, — и `safe_text` для показа.

6. Нефункциональные требования

Триаж должен укладываться в 300 мс, проверка ответа — в 2 с, и всё это на серверах заказчика без обращений к внешним сервисам.

Требование	Значение
Развёртывание	контейнеры Docker на серверах заказчика в РК; работа без доступа в интернет
Оборудование	триаж — на CPU; проверка — не более одного GPU с 24 ГБ памяти; исполнитель указывает точную конфигурацию в отклике
Задержка (p95)	`/v1/triage` ⩽ 300 мс, `/v1/verify` ⩽ 2 с на ответ до 1 500 символов
Пропускная способность	не менее 20 одновременных запросов на одном экземпляре без роста p95 сверх указанного
Отказоустойчивость	при недоступности проверки сервис возвращает осторожное решение (безопасный шаблон), а не «показать как есть»; триаж работает независимо от проверки
Языки	казахский (кириллица и латиница), русский, английский и их смесь; архитектура допускает добавление языка без переделки сервиса
Персональные данные	сервис не принимает имена, телефоны и идентификаторы пользователей; тексты не хранятся по умолчанию (ФТ-18); соответствие Закону РК «О персональных данных и их защите»
Безопасность	TLS, токены с ограниченными правами, журнал действий в редакторе базы знаний; сторонние модели и библиотеки — только с лицензией, допускающей коммерческое использование
Воспроизводимость	скрипты обучения и оценки запускаются одной командой и дают те же метрики (±0,5 п.п.) на оборудовании заказчика

Требование	Значение
Документация	руководство администратора, руководство врача-редактора, описание API, отчёт о НИОКР на русском языке

7. Данные и тест-набор

Реальные сообщения пользователей приложения исполнителю не передаются. Обучающие и тестовые данные исполнитель собирает сам: синтетические обращения, перефразирование открытых медицинских материалов и записи, составленные врачами и носителями языка по заданным сценариям.

7.1 Состав закрытого тест-набора

Тест-набор — 3 000 обращений и 1 000 ответов модели с размеченными утверждениями. Он создаётся на этапе 1 и до приёмки хранится только у заказчика.

Срез	Доля обращений
Казахский (кириллица)	30 %
Казахский (латиница)	10 %
Русский	30 %
Смесь казахского и русского в одном сообщении	20 %
Английский	10 %

Уровни срочности: «экстренно» — не менее 20 %, «срочно к врачу» — 25 %, «планово» — 30 %, «не медицинский» — 25 %. Не менее 15 % обращений — «ловушки»: отрицание («кровотечения нет»), прошлое («в прошлую беременность было»), третье лицо («у подруги»), гипотетика («а если вдруг…»).

В 1 000 ответов модели заранее заложены ошибки трёх типов: противоречие протоколу, неподтверждённое утверждение и запрещённый тип (диагноз, препарат, дозировка). Доля ответов с ошибкой — около 40 %.

7.2 Разметка

- Каждое обращение и каждый ответ размечают независимо два врача: акушер-гинеколог и педиатр — по профилю обращения.
- Расхождения решает третий врач-арбитр.
- Согласованность разметчиков (каппа Коэна) по уровню срочности — не ниже 0,8; при меньшем значении инструкция разметки уточняется и спорная часть размечается заново.
- Инструкция разметки — часть результатов работ и передаётся заказчику.

7.3 Обучающие данные

Объём и состав обучающих данных определяет исполнитель. Обращения из закрытого тест-набора и их перефразирование в обучении использовать запрещено; исполнитель подтверждает это при приёмке и передаёт скрипт проверки на пересечение.

8. Методика и критерии приёмки

Работа принимается по цифрам на закрытом тест-наборе (раздел 7), который исполнитель до приёмки не видит. Главное требование — не пропускать экстренные состояния; лишние срабатывания допускаются в ограниченной доле.

8.1 Метрики

Метрика	Порог
Полнота уровня «экстренно»	⩾ 98 % на всём наборе и ⩾ 96 % в каждом языковом срезе из п. 7.1
Недооценка на два уровня и более («экстренно» → «планово» или «не медицинский»)	⩽ 0,5 % экстренных обращений
Ложное «экстренно» среди обращений уровней «планово» и «не медицинский»	⩽ 10 %
Точность на «ловушках» (отрицание, прошлое, третье лицо, гипотетика)	⩾ 90 %
Средняя F1 по четырём уровням триажа	⩾ 0,85

Метрика	Порог
Опасный ответ пропущен без изменений (ответ с противоречием протоколу или запрещённым типом получил решение `show`)	⩽ 1 % таких ответов
Ложная замена корректного ответа (решение `replace`)	⩽ 5 % корректных ответов
Лишняя правка корректного ответа (решение `show_edited`)	⩽ 15 % корректных ответов
Верная ссылка на положение базы знаний у подтверждённых утверждений	⩾ 95 %
Задержка и пропускная способность	по разделу 6

Для каждой метрики в протоколе указывается 95-процентный доверительный интервал. Для двух первых строк таблицы порог сравнивается с точечной оценкой, а нижняя граница интервала полноты не должна опускаться ниже 95 %.

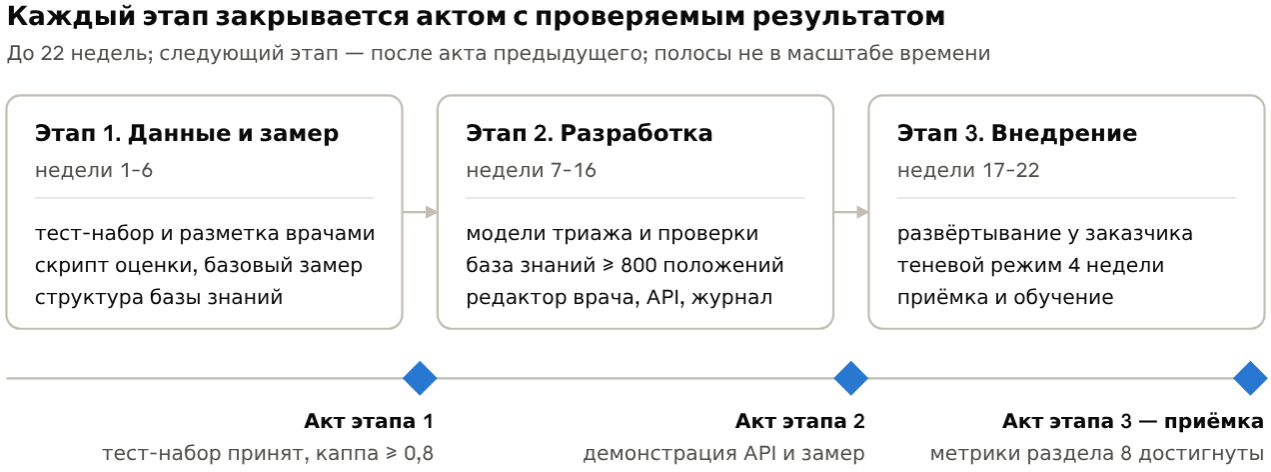
8.2 Порядок приёмки

1. Исполнитель разворачивает сервис на серверах заказчика по своей документации; заказчик проверяет, что сервис работает без доступа в интернет.
2. Заказчик прогоняет закрытый тест-набор через API и считает метрики скриптом оценки, переданным исполнителем на этапе 1.
3. Нагрузочный тест: задержка и пропускная способность на оборудовании из отклика исполнителя.
4. Врачи заказчика выборочно проверяют не менее 200 положений базы знаний на соответствие источнику и корректность перевода на три языка; допустимо не более 2 % замечаний и ни одного положения, противоречащего источнику.
5. Опытная эксплуатация в теневом режиме 4 недели: решения сервиса пишутся в журнал, но не меняют ответы пользователям; их разбирают врачи заказчика, данные исполнителю не передаются. Итог — обезличенная сводка расхождений с врачами.

6. Протокол подписывают обе стороны. Если хотя бы одна метрика не достигнута, исполнитель дорабатывает сервис и имеет право на одну повторную приёмку в течение 4 недель. Отчёт о повторной приёмке показывает только значения метрик, без самих обращений.

9. Этапы и сроки

Общий срок — до 22 недель от подписания договора. Работы идут в три этапа; каждый закрывается актом, и следующий этап начинается после подписания акта предыдущего. Исполнитель может предложить другое деление срока внутри 22 недель, сохранив три акта.



Этапы работ · 3 этапа, 3 акта, до 22 недель

Этап	Недели	Результат	Условие закрытия
1. Данные и замер	1–6	закрытый тест-набор (раздел 7), инструкция разметки, скрипт оценки; замер универсальной языковой модели на тест-наборе как точка отсчёта; структура базы знаний и список источников	тест-набор передан заказчику, каппа Коэна ≥ 0,8

Этап	Недели	Результат	Условие закрытия
2. Разработка	7–16	модели триажа и проверки, база знаний ≥ 800 положений, редактор врача, API по разделу 5, журнал и очередь разбора	демонстрация API на открытых примерах, отчёт о внутреннем замере
3. Внедрение и приёмка	17–22	развёртывание на серверах заказчика, нагрузочный тест, теневой режим 4 недели, документация, обучение	протокол приёмки по разделу 8

10. Результаты работ и права

10.1 Что передаётся заказчику

- Исходный код сервиса, редактора базы знаний и скриптов обучения и оценки.
- Веса всех обученных моделей и контейнеры для развёртывания.
- База знаний (не менее 800 положений на трёх языках, со ссылками на источники) и список сигналов триажа.
- Обучающие данные, закрытый тест-набор и инструкция разметки.
- Описание API в формате OpenAPI, руководства администратора и врача-редактора.
- Отчёт о НИОКР: подходы, результаты по каждому языку, известные ограничения и типичные ошибки.
- Обучение сотрудников заказчика: два занятия — для разработчиков и для врачей-редакторов.
- Гарантийное сопровождение 3 месяца после приёмки: исправление ошибок и одно дообучение по итогам опытной эксплуатации.

10.2 Что не входит в работы

- Постановка диагноза, назначение лечения и дистанционное консультирование — модуль только направляет к врачу и не даёт непроверенных советов.
- Регистрация модуля как медицинского изделия.

- Доступ к данным пользователей приложения, даже обезличенным.
- Изменение модели ответов заказчика и интерфейса приложения — их дорабатывает заказчик.
- Закупка и аренда серверов.

10.3 Права

Исключительные имущественные права на код, веса моделей, базу знаний, данные и документацию переходят к заказчику. Использование открытых моделей и библиотек допускается, если их лицензии разрешают коммерческое использование и перечислены в отчёте. Публикация результатов исследования возможна по согласованию с заказчиком.

11. Требования к отклику

Отклик исполнителя должен содержать:

1. Команду: специалисты по обработке естественного языка и практикующие врачи (акушер-гинеколог и педиатр), кто из них владеет казахским языком.
2. Подход к триажу и к проверке утверждений: какие модели и методы, как учитываются отрицание, срок беременности и возраст ребёнка.
3. Опыт работы с казахским языком и медицинскими текстами: предыдущие проекты и достигнутые метрики.
4. План сбора и разметки данных без использования реальных персональных данных.
5. Требования к оборудованию и ожидаемая задержка на нём.
6. Календарный план по этапам раздела 9, стоимость по этапам и условия гарантии.

Отклики оцениваются по критериям: квалификация команды и наличие врачей — 30 %, проработка технического подхода — 30 %, опыт с казахским языком — 20 %, стоимость и сроки — 20 %. Заказчик может пригласить исполнителей на техническую встречу и предложить небольшое пробное задание на открытых примерах.