ТЕХНИЧЕСКОЕ ЗАДАНИЕ (MVP)

Проект: Daryn Online

Интеллектуальная система оценивания развёрнутых ответов и эссе на казахском и русском языках по критериям преподавателя на основе искусственного интеллекта

Наименование	Модуль ИИ-проверки письменных работ и автоматической обратной связи (MVP)
Компания-заявитель (заказчик)	ТОО «Daryn.online» — владелец и оператор образовательной платформы Daryn Online
Контактное лицо	Мәдениет Дидар — ответственное лицо заказчика
Пользователи MVP	Преподаватели Daryn Online; администратор модуля
Данные	Задания и рубрики Daryn Online (10–15 заданий, DOCX/XLSX), синтетические работы исполнителя (≥ 1 500, каз./рус.), контрольный набор с оценками 2 экспертов (300 работ, XLSX), открытые корпуса эссе (ASAP-AES, ELLIPSE). Задания и рубрики передаёт заказчик в течение 5 рабочих дней (облачная папка или Git-репозиторий). Реальные работы с персональными данными исполнителю не передаются (раздел 3.1)
Срок разработки MVP	6 недель: разработка MVP — 4 недели, пилот — 2 недели (разделы 3.4 и 10)
Бюджет (диапазон)	от 900 000 до 1 500 000 тенге (договорная, в пределах диапазона), раздел 11
Форма вознаграждения	Договорная: от 900 000 до 1 500 000 тенге. Оплата поэтапно — 30% аванс, 40% после демонстрации промежуточного результата, 30% после приёмки MVP.
Примечание	Только резиденты РК. К заявке приложить описание решения и архитектуры, состав команды, портфолио аналогичных проектов (ИИ/ML, NLP), календарный план и смету. Источники, формат и объём данных, способ доступа и интеграции — в разделах 3.1–3.2; границы пилота — в разделе 3.4. Реальные персональные данные исполнителю не передаются. Исключительные права на код, модели и документацию переходят заказчику. Оплата LLM-API или GPU-сервера и хостинга — за счёт заказчика.
Дата	23.11.2026 (редакция 1)

1. Проблема

Проверка эссе и развёрнутых ответов занимает у преподавателя основную часть времени: на один класс уходит несколько часов. Обратная связь приходит учащемуся с опозданием на несколько дней, когда тема уже пройдена. Оценки разных преподавателей по одной работе заметно расходятся, а комментарии часто сводятся к итоговому баллу без объяснения ошибок. На казахском языке инструментов автоматической проверки практически нет.

2. Цель MVP

За минимальный срок создать рабочий модуль, который по критериям (рубрике) преподавателя оценивает эссе и развёрнутые ответы на казахском и русском языках, предлагает балл по каждому критерию и формирует понятный учащемуся комментарий, — и проверить его в двухнедельном пилоте с 5–10 преподавателями заказчика. MVP и пилот должны подтвердить, что оценки модуля согласуются с экспертами не хуже, чем оценки двух экспертов между собой, и что время проверки сокращается не менее чем на 50%.

Модуль является самостоятельным продуктом и не выставляет итоговую оценку: предложенный балл подтверждает или исправляет преподаватель.

3. Границы MVP

Входит в MVP	Не входит (следующие версии)
Генератор синтетических работ и загрузка работ в CSV/DOCX/TXT	Подключение реальных работ учащихся Daryn Online
Конструктор рубрики: до 6 критериев, шкала по каждому	Распознавание рукописного текста (OCR)
Балл по каждому критерию и итоговый балл	Проверка устных ответов (речь)
Комментарий: 2 сильные стороны, 3 ошибки с цитатой из текста	Детекция ИИ-сгенерированного текста и антиплагиат
Языки работ: казахский и русский	Английский и другие языки
Панель преподавателя: очередь работ, подтверждение/правка балла	Панель учащегося и родительский доступ
Экспорт результатов в CSV	Отчёты PDF/XLSX, аналитика по классу
REST API модуля	SSO и встраивание в интерфейс Daryn Online

3.1. Источники данных

Все материалы заказчика передаются в течение 5 рабочих дней после подписания договора через облачную папку или Git-репозиторий заказчика.

Источник	Формат	Объём	Способ доступа в MVP	Кто предоставляет
Типовые задания Daryn Online (эссе и развёрнутые ответы) по 2 предметам: казахский язык и литература, русский язык и литература	DOCX / XLSX	10–15 заданий, 7–11 классы	Файлы в облачной папке / Git-репозитории	Заказчик (методист)
Рубрики оценивания к заданиям	XLSX по шаблону ис полнителя	До 6 критериев на задание, шкала и описание уровней	Тот же канал; шаблон передаёт исполнитель на неделе 1	Заказчик
Синтетические работы	CSV / TXT	≥ 1 500 работ (каз./рус. поровну), до 500 слов	Создаются исполнителем	Исполнитель
Контрольный набор с экспертными оценками	XLSX	300 работ × 2 независимые оценки	Разметка в панели или XLSX на неделе 2	Заказчик (2 эксперта)
Открытые корпуса эссе с оценками (для отработки методики)	CSV	ASAP-AES (~13 000 эссе), ELLIPSE (~6 500 эссе)	Загрузка из открытых источников; проверка лицензии	Исполнитель
Работы для пилота	DOCX / TXT	≥ 200 работ (≥ 100 на каждом языке)	Загружают преподаватели в панель на сервере заказчика; без ФИО и ИИН	Заказчик (учащиеся-добровольцы с согласия законных представителей)

Документация. Вместе с заданиями заказчик передаёт описание формата: перечень заданий (предмет, класс, язык, тип ответа, максимальный балл) и рубрики по шаблону исполнителя.

Не используются: ФИО, ИИН, контакты и иные персональные данные учащихся. Работы пилота хранятся и обрабатываются только на сервере заказчика.

3.2. Информационные системы и интеграции

Система	Назначение	Интеграция в MVP	Кто обеспечивает
Платформа Daryn Online	Источник заданий и рубрик	Файловый обмен (DOCX/XLSX/CSV); встраивание в интерфейс платформы — следующая версия	Заказчик
Облачная папка или Git-репозиторий заказчика	Передача заданий, рубрик, разметки	Структура папок и шаблоны от исполнителя	Заказчик
Модуль проверки (REST API)	Отправка работы на проверку, получение баллов и комментария	FastAPI, документация OpenAPI, авторизация по ключу	Исполнитель
Веб-панель преподавателя	Очередь работ, подтверждение и правка балла	Браузер, вход по логину и паролю	Исполнитель
Языковая модель (LLM)	Оценивание и комментарий	Вариант А: внешний LLM-API (ключ на аккаунт заказчика). Вариант Б: открытая модель (ISSAI KazLLM, Qwen) на сервере с GPU. Выбор — на неделе 1	Заказчик оплачивает; исполнитель настраивает
Проверка орфографии и грамматики	Поиск языковых ошибок	Открытые словари каз./рус. (Hunspell или аналог) в составе модуля	Исполнитель
Хостинг в РК (Alem Plus, PS Cloud или иной)	Размещение модуля и данных пилота	VPS: 4 vCPU, 8 ГБ RAM, 100 ГБ SSD (вариант А) или сервер с GPU 24 ГБ (вариант Б)	Заказчик предоставляет доступ

3.3. Обязательства заказчика

- в течение 5 рабочих дней передать 10–15 заданий и рубрики по шаблону исполнителя;
- предоставить доступ к хостингу и репозиторию, ключ LLM-API (при варианте А);
- выделить 2 экспертов для разметки контрольного набора (300 работ) на неделе 2;
- обеспечить пилот: 5–10 преподавателей и не менее 200 обезличенных работ;
- назначить ответственное лицо для оперативных согласований (Мәдениет Дидар).

3.4. Пилот (MVP): границы

Параметр	Значение
Объект пилота	Веб-панель проверки, развёрнутая на сервере заказчика; оценивание по рубрике и комментарии
Предметы и типы работ	Казахский язык и литература, русский язык и литература; эссе и развёрнутые ответы до 500 слов; 7–11 классы
Участники	5–10 преподавателей заказчика
Объём	Контрольный набор — 300 работ; в пилоте — не менее 200 работ (не менее 100 на каждом языке)
Замер времени	Каждый преподаватель проверяет половину работ вручную, половину — с модулем; время фиксируется в панели
Сроки	Разработка MVP — недели 1–4; пилот — недели 5–6; итоговый отчёт — конец недели 6
Критерии оценки	Метрики раздела 9 на контрольном наборе (неделя 4) и в пилоте (недели 5–6)
Не входит в пилот	Другие предметы, рукописные и устные работы, выставление итоговых оценок, встраивание в платформу
Результат	Отчёт о пилоте с метриками и рекомендацией о разработке полной версии

4. Пользовательский сценарий

1	Преподаватель создаёт задание и рубрику (критерии и шкалу) или выбирает готовую.
2	Администратор загружает работы (CSV/DOCX/TXT) и запускает проверку.
3	Система показывает очередь работ с предложенными баллами по критериям и уровнем уверенности.
4	Преподаватель открывает работу: текст с подсветкой ошибок, баллы и комментарий; подтверждает или правит.
5	Преподаватель выгружает результаты в CSV; итоговую оценку всегда утверждает преподаватель.

5. Функциональные требования

№	Функция	Требование MVP
1	Генерация данных	Не менее 1 500 синтетических работ на 2 языках, 5 уровней качества
2	Загрузка работ	CSV, DOCX, TXT; проверка формата; сообщение об ошибке
3	Рубрика	До 6 критериев, шкала и описание уровней по каждому критерию
4	Оценивание	Балл по каждому критерию и итоговый балл; уровень уверенности
5	Обратная связь	2 сильные стороны и 3 ошибки с цитатой из текста и советом, на языке работы
6	Проверка темы	Пометка «не по теме» / «слишком короткий ответ»
7	Панель преподавателя	Очередь работ, фильтры, подтверждение и правка балла, история правок
8	Экспорт	Выгрузка результатов в CSV
9	API	Отправка работы на проверку, получение результата, авторизация по ключу
10	Отчёт о качестве	Согласованность с экспертами, сравнение казахского и русского

6. Контроль качества оценивания

- **Уверенность:** при низкой уверенности работа помечается «требует ручной проверки».
- **Обоснованность:** каждая ошибка подтверждается цитатой, которая действительно есть в тексте.
- **Согласованность:** итоговый балл равен сумме баллов по критериям и не выходит за шкалу.
- **Стабильность:** повторная проверка той же работы даёт балл с отклонением не более 1.
- **Отсутствие перекоса:** качество проверяется отдельно по языкам; разница QWK не более 0,10.

7. Технологическое решение

Компонент	Решение для MVP
Генерация данных	Python; LLM-генерация работ по сценариям качества с ручной выборочной проверкой
Модели	LLM с промптами по рубрике + калибровочная модель (регрессия по признакам текста) для балла
Языковой анализ	Орфография и грамматика для казахского и русского языков
API	FastAPI (или согласованный), документация OpenAPI
Интерфейс	Веб-панель преподавателя (React, Vue или серверные шаблоны)
Хранение	PostgreSQL: работы, рубрики, оценки, правки преподавателя
Развертывание	Docker, docker-compose; GPU 24 ГБ при локальной LLM

Безопасность: вход по логину и паролю; роли преподаватель и администратор; доступ к API по ключу; HTTPS; секреты вне кода; в LLM не передаются персональные данные. Оценка носит рекомендательный характер.

8. Нефункциональные требования

- Проверка одной работы (до 500 слов) — не более 30 секунд; 100 работ — не более 15 минут.
- Время ответа панели — не более 2 секунд.
- Одновременная работа не менее 10 преподавателей.
- Данные не теряются при перезапуске контейнеров; журнал запусков проверки и правок.

9. Критерии приёмки MVP и оценки пилота

Метрика	Где измеряется	Целевое значение
QWK между модулем и средней оценкой экспертов	Контрольный набор (300 работ)	≥ 0,70
Доля работ с отклонением балла не более 1	Контрольный набор	≥ 85%
Согласованность модуля не ниже согласованности двух экспертов	Контрольный набор	Да
Разница QWK между казахским и русским языками	Контрольный набор	≤ 0,10
Доля ошибок в комментариях, подтверждённых цитатой	Контрольный набор	≥ 95%
Доля баллов, принятых преподавателем без правки или с правкой ≤ 1	Пилот (≥ 200 работ)	≥ 80%
Сокращение времени проверки	Пилот (замер в панели)	≥ 50%
Оценка полезности комментариев преподавателями	Опрос участников пилота	≥ 4 из 5
Персональные данные в наборах исполнителя	Проверка наборов	Отсутствуют

10. Этапы разработки и пилота

Этап	Срок	Результат
Подготовка и данные	Неделя 1	Получение заданий и рубрик, выбор LLM (А/Б), генерация синтетических работ
Модели	Недели 2–3	Оценивание по рубрике, комментарии, разметка контрольного набора экспертами
Панель и API	Недели 2–3 (параллельно)	Веб-панель, API, развертывание в Docker
Приёмочные испытания	Неделя 4	Проверка метрик на контрольном наборе, отчёт о качестве
Пилот	Недели 5–6	Работа 5–10 преподавателей, ≥ 200 работ, замер времени, отчёт о пилоте

11. Бюджет и порядок оплаты

Ориентировочный бюджет MVP и пилота: **от 900 000 до 1 500 000 тенге**. Форма вознаграждения — договорная. В сумму входят работы исполнителя по всем этапам, включая сопровождение пилота. Оплата LLM-API или GPU-сервера и хостинга — за счёт заказчика.

Этап	Доля	Сумма, тенге (от – до)
Подготовка и данные	15%	135 000 – 225 000

Этап	Доля	Сумма, тенге (от – до)
Модели	35%	315 000 – 525 000
Панель и API	25%	225 000 – 375 000
Приёмочные испытания	10%	90 000 – 150 000
Пилот и отчёт	15%	135 000 – 225 000
Итого	**100%**	**900 000 – 1 500 000**

Оплата поэтапная: 30% — аванс после подписания договора; 40% — после приёмочных испытаний (неделя 4); 30% — после завершения пилота и подписания акта по критериям раздела 9.