ТЗ: модуль распознавания рукописных решений по математике, физике, химии и геометрии

Nov 25, 2026 · @Bekda

1. Общие сведения

Нужен сервис, который по фотографии рукописного решения задачи по математике, физике, химии или геометрии возвращает его цифровую версию: текст, формулы в LaTeX, химические структуры в SMILES и выделенные рисунки с подписями.

Ученики решают задачи на бумаге и фотографируют страницы на смартфон. В одном решении смешаны текст на русском или казахском, многострочные выкладки, химические уравнения, структурные формулы молекул, чертежи, графики и схемы. Обычный OCR рассчитан на печатный текст и такие страницы не разбирает.

Эта задача — MVP: распознавание текста, формул, химических уравнений и простых структурных формул, а также обнаружение рисунков с их типом и подписями, около 10 недель. Полная векторизация чертежей, оцифровка графиков и электрических схем — план развития вне этой задачи.

Модуль не оценивает правильность решения — он только переводит страницу в цифровой вид. Модуль поставляется как сервис с API и разворачивается в арендуемом заказчиком GPU-облаке на территории Казахстана.

2. Термины

Термин	Определение
Решение	Одна или несколько страниц, на которых ученик от руки решил одну или несколько задач
Страница	Одно изображение: фото или скан листа

Термин	Определение
Блок	Область страницы одного типа (текст, формула, химическое уравнение, структурная формула, рисунок, таблица, зачёркнутое) с рамкой и порядковым номером
Формула	Математическое или физическое выражение, выдаётся в LaTeX
Химическое уравнение	Формулы веществ и уравнения реакций, выдаются в нотации mhchem (`\ce{...}`)
Структурная формула	Рисунок молекулы с атомами и связями, выдаётся в SMILES и MOL
Рисунок	Чертёж, график или схема: геометрическая фигура, стереометрический чертёж, схема сил, электрическая или оптическая схема
Подпись на рисунке	Буква вершины, число, обозначение угла, длины или величины с координатами на рисунке
Уверенность	Оценка 0–1 у каждого блока; фрагменты ниже порога помечаются «неразборчиво»

3. Границы работ (MVP)

Эта задача — MVP: сервис переводит фото рукописного решения в структурированный документ с текстом, формулами, химией и выделенными рисунками.

Входит в MVP:

- предобработка фото: обрезка листа, выравнивание, подавление теней и клетки, оценка качества;
- разметка страницы на блоки и порядок чтения, номера задач и пунктов;
- рукописный текст на русском и казахском и формулы в LaTeX;
- химические формулы и уравнения в mhchem;
- структурные формулы школьных органических молекул (до 15 неводородных атомов) в SMILES и MOL;
- рисунки: рамка, тип, вырезанное изображение и подписи с координатами;
- уверенность и пометка неразборчивых и зачёркнутых мест;

- REST API, итоговый JSON и Markdown, HTML-страница сравнения для приёмки;
- сбор и разметка обучающих данных, дообучение открытых моделей, Docker-образы и передача кода, весов и данных.

Не входит:

- проверка правильности решения и выставление оценок;
- загрузка фото учеником и интерфейсы — обеспечивает заказчик;
- определение автора по почерку.

План развития (вне этой задачи и бюджета): векторизация геометрических чертежей (точки, отрезки, окружности, отношения параллельности и перпендикулярности), оцифровка графиков, электрические схемы в список элементов и соединений, стереохимия и сложные полициклы, рукописный ввод с планшета в реальном времени, другие предметы.

4. Входные данные

Модуль работает с бытовыми фото тетрадных страниц, снятыми смартфоном, а также со сканами и PDF.

Модуль устойчив к: перспективе и наклону до 30°, неровному и загнутому листу, теням и бликам, клетке и линейке тетради, синей, чёрной и цветной ручке и карандашу, исправлениям и зачёркиваниям, плотному почерку, записям на полях, развороту из двух страниц в одном кадре.

Параметр	Значение
Форматы	JPEG, PNG, HEIC, PDF (каждая страница — отдельное изображение)
Разрешение	от 2 Мп, целевое 8–12 Мп
Страниц в решении	1–20
Языки текста	русский и казахский (включая ә, ғ, қ, ң, ө, ұ, ү, һ, і); латиница и греческие буквы в формулах
Предметы	математика (алгебра и начала анализа), геометрия (планиметрия и стереометрия), физика, химия — школьный уровень
Доставка	ссылка на S3-совместимое хранилище заказчика или загрузка через API

5. Типы содержимого

В эту задачу входят шесть типов A (C01–C06); остальные пять — план развития: они приведены, чтобы архитектура их не исключала.

Код	Тип	Что распознаётся	Приоритет
C01	Текст	Рукописный текст на русском и казахском: пояснения, «Дано», «Найти», «Ответ»	A — MVP
C02	Формулы и выкладки	Дроби, степени, корни, индексы, системы, неравенства, модули, логарифмы, тригонометрия, пределы, производные, интегралы, векторы, единицы СИ; многострочные преобразования	A — MVP
C03	Химические формулы и уравнения	Формулы веществ, коэффициенты, индексы, заряды ионов, стрелки, условия над стрелкой, ионные уравнения	A — MVP
C04	Структурные формулы	Органические молекулы школьного курса до 15 неводородных атомов: цепи, циклы, бензольное кольцо, кратные связи, функциональные группы → SMILES и MOL	A — MVP
C05	Рисунки	Рамка, тип рисунка, вырезанное изображение, подписи с координатами	A — MVP
C06	Зачёркнутое и исправления	Зачёркнутое помечается и не попадает в основной текст; вставки ставятся на своё место	A — MVP
C07	Таблицы	Таблицы данных в решениях по физике и химии → Markdown-таблица	B — развитие

Код	Тип	Что распознаётся	Приоритет
C08	Векторизация чертежей	Точки, отрезки, окружности, углы и отношения между ними	B — развитие
C09	Оцифровка графиков	Оси, масштаб, точки и кривые	B — развитие
C10	Электрические схемы	Элементы цепи и соединения	B — развитие
C11	Сложная химия	Стереохимия, конденсированные полициклы, механизмы реакций	B — развитие

Типы рисунков для C05: геометрическая фигура, стереометрический чертёж, график функции, схема сил и векторов, электрическая схема, оптическая схема, прочее. A — входит в MVP; B — план развития, вне этой задачи и бюджета.

6. Функциональные требования

На входе — страницы решения, на выходе — структурированный документ: блоки с типом, координатами, содержимым и уверенностью.



конвейер распознавания · от фото до документа

Разметка направляет каждый блок в свою ветку; сборка возвращает их в порядок чтения и помечает сомнительные места.

1. **FR-1. Приём.** 1–20 страниц на решение; синхронный режим для одной страницы и асинхронный для пакета; ссылкой на хранилище или загрузкой.
2. **FR-2. Предобработка.** Поиск листа, выравнивание перспективы и наклона, подавление теней, клетки и линовки, разделение разворота. Если страница непригодна (размыто, темно, обрезано), модуль возвращает отказ с причиной, а не пустой результат.
3. **FR-3. Разметка страницы.** Блоки типов C01–C06 с рамками, порядок чтения, номера задач и пунктов (1, 2, а), б)).
4. **FR-4. Текст и формулы.** Русский и казахский текст, формулы в LaTeX, многострочные выкладки как выровненные строки. Итоговый ответ выделяется отдельным блоком, если он отмечен в решении (подчёркнут, обведён или подписан «Ответ»).
5. **FR-5. Химия.** Формулы и уравнения в mhchem; структурные формулы в SMILES и MOL с проверкой валентностей. Невалидная структура помечается, изображение блока сохраняется.
6. **FR-6. Рисунки.** Рамка, тип из списка раздела 5, вырезанное изображение в хранилище заказчика, подписи с координатами относительно рисунка.
7. **FR-7. Уверенность.** У каждого блока — уверенность 0–1; фрагменты ниже порога помечаются «неразборчиво», модуль не подставляет догадку.
8. **FR-8. Сборка.** Итоговый Markdown решения с LaTeX, mhchem, SMILES и ссылками на изображения рисунков в порядке чтения, плюс JSON со структурой и координатами.
9. **FR-9. Воспроизводимость.** Повторный прогон той же страницы той же версией модели даёт тот же результат.
10. **FR-10. Страница сравнения.** HTML-страница: изображение с наложенными блоками и распознанный результат рядом — для приёмки и разбора ошибок.

7. Формат результата и API

Модуль отдаёт REST API со спецификацией OpenAPI 3; результат по решению — JSON со структурой и Markdown для отображения.

Метод	Путь	Назначение
POST	/v1/recognize	Распознать одну страницу синхронно
POST	/v1/jobs	Создать задание на решение из нескольких страниц

Метод	Путь	Назначение
GET	/v1/jobs/{id}	Статус задания
GET	/v1/jobs/{id}/result	Результат: JSON со страницами и блоками
GET	/v1/jobs/{id}/result.md	Итоговый Markdown решения
GET	/v1/jobs/{id}/review.html	Страница сравнения для приёмки

Вебхуки `job.done` и `job.failed` с подписью HMAC; идемпотентность по заголовку `Idempotency-Key`. Координаты рамок — доли страницы [x, y, ширина, высота].

Пример результата (сокращён):

```
{
  "jobId": "j-000451",
  "modelVersion": "1.0.0",
  "pages": [{
    "index": 1,
    "quality": { "usable": true },
    "blocks": [
      { "id": "b1", "order": 1, "type": "text", "bbox": [0.08, 0.06, 0.62,
0.04],
        "content": "Дано: $m = 2$ кг, $v_0 = 10$ м/с", "confidence": 0.94 },
      { "id": "b2", "order": 2, "type": "formula", "bbox": [0.08, 0.12, 0.70,
0.09],
        "latex": "E_k = \\frac{m v_0^2}{2} = 100\\,\\text{Дж}", "confidence":
0.88 },
      { "id": "b3", "order": 3, "type": "chem_structure", "bbox": [0.10,
0.30, 0.35, 0.18],
        "smiles": "CC(=O)O", "image": "s3://ocr/j-000451/b3.png",
"confidence": 0.81 },
      { "id": "b4", "order": 4, "type": "drawing", "drawingType":
"geometry_plane",
        "bbox": [0.55, 0.30, 0.38, 0.30], "image": "s3://ocr/j-
000451/b4.png",
        "labels": [{ "text": "A", "xy": [0.12, 0.91] }, { "text": "60°",
"xy": [0.20, 0.80] }],
        "confidence": 0.90 },
      { "id": "b5", "order": 5, "type": "answer", "latex":
"100\\,\\text{Дж}", "confidence": 0.93 }
    ]
  }]
}
```

8. Требования к качеству (KPI)

Главное требование — не выдавать уверенный неверный результат: сомнительное место должно быть помечено, а не угадано.

KPI считаются на закрытом тестовом наборе заказчика (раздел 10) по каждому предмету отдельно; LaTeX и SMILES сравниваются после нормализации (каноническая форма).

KPI	Как считается	Цель
Ошибка в тексте (C01)	Доля неверных символов (CER) в строках текста	⩽ 8%
Формулы без ошибок (C02)	Доля формул, чей LaTeX совпал с разметкой	⩾ 65%
Ошибка в формулах (C02)	Нормированное расстояние редактирования по токенам LaTeX	⩽ 10%
Итоговые ответы	Доля верно распознанных ответов (число, выражение, единица)	⩾ 90%
Химические уравнения (C03)	Ошибка по токенам mhchem	⩽ 10%
Структурные формулы (C04)	Доля молекул с совпавшим каноническим SMILES	⩾ 70%
Рисунки: обнаружение (C05)	Полнота обнаружения областей при перекрытии с разметкой ⩾ 70%	⩾ 0,95
Рисунки: тип (C05)	Доля верно определённых типов	⩾ 0,85
Подписи на рисунках (C05)	Полнота распознанных подписей (буквы, числа, углы, длины)	⩾ 0,85
Зачёркнутое (C06)	Доля зачёркнутых фрагментов, попавших в основной текст	⩽ 5%
Ложная уверенность	Доля ошибочных блоков среди помеченных как уверенные	⩽ 5%
Скорость	Время обработки одной страницы на одном GPU уровня NVIDIA L4, 95-й перцентиль	⩽ 20 с
Пропускная способность	Страниц в час на одном GPU-сервере при ⩾ 10 параллельных запросах	⩾ 300

KPI	Как считается	Цель
Разрыв русский / казахский	Разница CER между русским и казахским текстом	⩽ 3 п.п.

9. Среда и нефункциональные требования

Модуль работает в GPU-облаке на территории Казахстана, которое арендует заказчик; изображения и результаты не покидают это облако.

Область	Требование
Среда	Виртуальная машина в казахстанском GPU-облаке: Linux x86_64, NVIDIA GPU уровня L4 (24 ГБ) или выше, CUDA. Конфигурация согласуется на этапе 1, аренду оплачивает заказчик
Развёртывание	Docker-образы и docker compose; все модели работают локально, внешние AI-API и облачные OCR-сервисы не используются
Хранение	S3-совместимое хранилище для изображений и вырезанных рисунков, PostgreSQL для заданий и результатов
Скорость	По разделу 8: страница ⩽ 20 с, ⩾ 300 страниц в час на одном GPU-сервере; масштабирование — добавлением серверов
Надёжность	Падение обработчика не теряет задание, обработка продолжается после перезапуска
Безопасность	TLS; доступ к API по ключам с ротацией; журнал доступа к результатам
Наблюдаемость	JSON-логи, health-проба, базовые метрики (очередь, время на страницу, ошибки, доля неразборчивых блоков)
Документация	На русском; технические части допускаются на английском

10. Данные, персональные данные и лицензии

Обучающие данные собирает и размечает исполнитель, закрытый тестовый набор готовит заказчик; реальные страницы учеников обрабатываются только в облаке заказчика в РК.

Данные	Кто предоставляет	Объём и условия
Образцы для старта	Заказчик	30–50 обезличенных страниц реальных решений по четырём предметам в первую неделю — для проверки форматов и почерков
Обучающие страницы с разметкой	Исполнитель	Решения, написанные от руки участниками с письменного согласия, открытые датасеты с коммерческой лицензией, синтетика; объём — достаточный для KPI раздела 8; разметка по протоколу, согласованному на этапе 1
Закрытый тестовый набор	Заказчик	Около 400 страниц, по 100 на математику, физику, химию и геометрию, размечены по протоколу этапа 1; исполнителю не передаётся

Персональные данные

- Соответствие Закону РК «О персональных данных и их защите».
- Образцы передаются обезличенными: имена и другие идентификаторы на листе заказчик закрывает до передачи.
- Модуль не хранит изображения дольше обработки, срок хранения результатов настраивается; определение автора по почерку не выполняется.
- Участники сбора обучающих данных дают письменное согласие; согласия передаются заказчику вместе с данными.

Допустимые лицензии моделей, библиотек и датасетов

- Разрешены Apache-2.0, MIT, BSD и аналогичные, допускающие коммерческое использование в закрытом продукте.

- Не допускаются AGPL и GPL, веса с некоммерческой лицензией и датасеты «только для исследований» (к ним относится ряд известных наборов рукописных формул) — для обучения итоговых весов их использовать нельзя.
- Модели с собственными лицензиями допустимы, только если лицензия разрешает коммерческое использование без ограничений; перечень согласуется на этапе 1.
- Перечень зависимостей, моделей и датасетов с лицензиями (SBOM) — часть поставки.

11. Этапы, результаты и права

Работа идёт тремя этапами, около 10 недель; оплата — по актам приёмки этапов. После приёмки — 3 месяца гарантии на исправление дефектов.



этапы работ · 3 этапа, 3 гейта

Закрытый тестовый набор заказчик размечает параллельно с этапами 1–2, чтобы к этапу 3 он был готов.

Передаваемые результаты

- Исходный код сервиса и скриптов обучения с историей репозитория.
- Веса моделей, конфигурации обучения, собранный датасет с разметкой и согласиями участников.
- Docker-образы, docker compose, инструкция развёртывания с нуля.
- Спецификация OpenAPI и документация интеграции.
- Отчёт о KPI на закрытом наборе по каждому предмету и типу содержимого.
- SBOM и перечень лицензий.

Права. Исключительные имущественные права на код, веса моделей и собранный датасет переходят заказчику по каждому этапу с момента подписания акта и оплаты этапа. Код и веса передаются в репозиторий заказчика не позднее 5 рабочих дней после акта. Исполнитель не использует данные заказчика вне проекта.

12. Порядок приёмки

Модуль принимается по KPI раздела 8 на закрытом тестовом наборе заказчика, который исполнителю не передаётся. Прогон идёт контейнером исполнителя в облаке заказчика; метрики считает скрипт оценки, согласованный на этапе 1.

1. Развёртывание с нуля по инструкции на чистой виртуальной машине заказчика — не дольше одного рабочего дня.
2. Прогон на закрытом наборе: все KPI раздела 8 по каждому предмету.
3. Проверка скорости: 10 параллельных запросов, не менее 300 страниц за час на одном GPU-сервере.
4. Функциональная проверка FR-1…FR-10 и контрактные тесты API по OpenAPI.
5. Безопасность: нет исходящих соединений во внешние сервисы, журнал доступа ведётся.
6. Комплектность поставки по разделу 11 и лицензионная чистота по SBOM.

13. Требования к исполнителю

Нужна команда с опытом продакшн-систем распознавания рукописного текста или формул.

- Не менее одного внедрённого проекта OCR рукописного текста, формул или документов — с контактами заказчика.
- В команде — ML-инженер по компьютерному зрению или OCR и бэкенд-разработчик; желателен опыт с химическими структурами (SMILES, RDKit).
- Опыт работы с персональными данными и готовность к NDA.
- Резидентство РК.