Техническое задание

LLM-матчинг профиля кандидата и вакансий для AI-платформы автоматизации поиска работы
Заказчик: ТОО «Global Career Company» (участник Astana Hub)

1. Контекст

Компания разрабатывает AI-платформу, которая автоматизирует поиск работы: собирает и обогащает вакансии, подбирает релевантные под профиль кандидата, автоматизирует отклики и аутрич. Платформа работает в двух форматах: self-serve кабинет пользователя и рабочий кабинет команды рисерч-аналитиков. В обработке около 2,7 млн вакансий. Текущий подбор построен на поисковых гипотезах, фильтрах и ключевых словах.

2. Цель

Разработать и внедрить модуль семантического матчинга на базе LLM, который оценивает соответствие вакансии профилю кандидата, выдаёт оценку релевантности с объяснением и сокращает ручную проверку вакансий.

3. Входные данные

- Профиль кандидата: целевые роли и грейд, стек и ключевые навыки, домен/индустрии, языки, локация и готовность к релокации, формат работы, ограничения (стоп-компании, стоп-слова), поисковые гипотезы.
- Вакансия: название, текст описания (RU/EN), компания, локация, формат работы, требования, зарплатная вилка (если есть), источник и дата публикации.

4. Результат работы модуля

- Оценка релевантности вакансии профилю по шкале 0–100.
- Разбивка по критериям: роль, грейд, стек, домен, язык, локация, формат работы.
- Краткое объяснение для пользователя (1–3 предложения): почему вакансия подходит или не подходит.
- Флаги жёстких несоответствий (например, обязательная релокация, язык, грейд) — вакансия исключается из выдачи.
- Нормализованные атрибуты вакансии: роль, грейд, стек — для повторного использования в поиске.

5. Функциональные требования

- Корректная работа с вакансиями на русском и английском языках.
- Распознавание синонимичных ролей и названий (например, Product Owner / Product Manager / Head of Product).
- Определение грейда по тексту вакансии, даже если он не указан явно.
- Учёт обратной связи пользователя (подходит / не подходит) для дообучения или калибровки оценки.
- Пакетная обработка новых вакансий и пересчёт оценок при изменении профиля.

6. Нефункциональные требования

- Производительность: обработка текущего потока вакансий; ориентир — оценка одной пары «профиль–вакансия» не дольше 2 с, пакетная обработка не менее 10 000 пар в час.

- Стоимость: прогнозируемая стоимость обработки в пересчёте на 1 000 пар; допускается каскад (дешёвый предфильтр + LLM для финального ранжирования).
- Интеграция: REST API или очередь сообщений; документация по API.
- Хранение и защита персональных данных кандидатов; возможность развёртывания на инфраструктуре заказчика или использования LLM через API с обезличиванием данных.

7. Критерии приёмки

- Доля релевантных вакансий в подборке — не ниже 80% по оценке пользователей на тестовой выборке.
- Сокращение ручной фильтрации вакансий не менее чем в 2 раза.
- Объяснение есть у 100% оценок и понятно пользователю без технической подготовки.
- Модуль работает на текущем потоке вакансий без существенного роста инфраструктурных затрат.

8. Формат взаимодействия и этапы

Этап	Содержание	Результат
1. Анализ	Изучение данных и текущего подбора, разметка тестовой выборки	Метрики и тестовый датасет
2. Прототип	Модель матчинга на тестовой выборке	Отчёт по качеству
3. Пилот	Интеграция с платформой на части пользователей	Сравнение с текущим подбором
4. Внедрение	Полное подключение, документация, передача	Рабочий модуль и API

Вознаграждение: условия сотрудничества и вознаграждение обсуждаются индивидуально после рассмотрения предложений. Заказчик оставляет за собой право не выбирать исполнителя. В предложении просим указать подход, опыт аналогичных проектов, сроки по этапам и ориентировочную стоимость.