

ML LLM инженер по обработке документов

от 170 000р

- Требуемый опыт работы: От 3 до 6 лет
- Занятость: Частичная
- График: Удаленная работа
- Образование: Высшее

О вакансии

TERRIXA — интеллектуальная система для работы с техническими документами и производственными данными. Ищем ML/LLM-инженера, который будет развивать подготовку данных, классификацию запросов и оценку качества ответов, а также участвовать в тестировании сквозных сценариев анализа документов и отчётов по скважинам.

Чем предстоит заниматься

- Формировать, размечать и актуализировать наборы пользовательских запросов и эталонных ответов для обучения и тестирования.
- Поддерживать сбалансированные выборки по типам задач: общий поиск по документам, точные значения из таблиц, графовые связи, расчёты и вызов агентов.
- Развивать и оценивать классификацию намерений пользователя и маршрутизацию запросов между компонентами системы.
- Проверять качество ответов на технических документах и суточных отчётах.
- Участвовать в проверке сводного анализа группы отчётов: по одной скважине за период и по нескольким скважинам за период.
- Оценивать полноту и корректность сводных результатов: выполненные работы, статусы, выявленные проблемы и динамика.
- Тестировать сквозной цикл обработки документов: загрузка, извлечение данных, запись в базы, выполнение запросов и получение результата, в том числе в Excel.
- Разбирать ошибки и случаи отсутствия результата, группировать причины и предлагать улучшения качества.
- Настраивать и оценивать локальный запуск моделей на GPU/CPU, включая Ollama или vLLM, параметры параллельной обработки и таймауты.
- Работать совместно с разработчиками парсера, backend-командой и тестировщиками; документировать эксперименты и результаты.

Что важно для этой работы

- Опыт работы ML-, NLP- или LLM-инженером.
- Уверенный Python и практические навыки обработки данных, в том числе с pandas.
- Опыт подготовки датасетов, разметки данных, оценки классификаторов и анализа ошибок.
- Понимание LLM/RAG-систем, поиска по документам и подходов к оценке качества ответов.
- Практический опыт работы с SQL и структурированными данными.
- Умение тестировать пайплайны обработки документов и находить причины некорректных результатов.
- Навыки работы в Linux-среде и чтения логов.
- Умение ясно фиксировать ход эксперимента, результаты и ограничения.

Будет преимуществом

- Опыт с Hugging Face Transformers, scikit-learn, PyTorch или аналогичными инструментами.
- Практика запуска LLM через Ollama или vLLM и настройки инференса на GPU.
- Опыт с векторными и графовыми базами данных.
- Опыт обработки PDF, DOCX, Excel и документов со сложными таблицами.
- Знакомство с нефтегазовой тематикой, буровыми процессами и суточными рапортами.
- Опыт нагрузочного тестирования на больших наборах документов.

Задачи на первые месяцы

- Разобраться в текущих наборах данных и сценариях классификации запросов.
- Согласовать воспроизводимый набор тестов и метрики качества для ключевых типов поиска и отчётных сценариев.
- Регулярно предоставлять команде разбор ошибок с примерами и предложениями по улучшению.

- Помочь стабилизировать оценку работы моделей в документном пайплайне и отчётных сценариях.

Адрес размещения вакансии:

Москва, Арбат, д. 24

Контактная информация:

- +7 (931) 521-07-00
- Людмила
- office@terrixa.com

Другие вакансии этой компании

Инженер технической поддержки ПО и оборудования ГТИ

от 150 000р

Устанавливать, настраивать и проверять ПО ГТИ, оборудование и каналы сбора и передачи данных. Консультировать пользователей по работе системы, разбирать обращения и фиксировать шаги диагностики. Выявлять причины сбоев в ПО, подключениях, оборудовании и передаче данных; собирать техническую информацию для передачи разработчикам. Проверять корректность поступления и отображения данных, включая временную и глубинную привязку; при наличии соответствующего контура — контролировать обмен по WITS0/WITSML. Собирать и монтировать электронные узлы и приборы для ГТИ по схемам, чертежам и технической документации. Выполнять ручную пайку электронных компонентов, проводов, разъёмов и кабельных сборок; проверять качество монтажа. Читать принципиальные электрические схемы, находить элементы и цепи, проверять соединения, полярность и монтаж. Проводить первичную диагностику электроники мультиметром и другим доступным измерительным оборудованием. Выполнять базовые операции с микроконтроллерами: подключать плату, загружать готовую прошивку, задавать параметры и проверять интерфейсы по инструкции. Проверять собранные устройства на стенде, участвовать в ремонте и подготовке оборудования к эксплуатации. Вести записи о диагностике, результатах проверок, выполненных работах и использованных компонентах.

деревня Крёкшино, Терминальный проезд, д. 5с13