



Немов Максим Юрьевич

Мужчина, 24 года, родился 4 августа 2000

+7 (987) 5056990 — предпочитаемый способ связи

aangers07@gmail.com

Другой сайт: <https://t.me/dizilx>

Проживает: Москва

Гражданство: Россия, есть разрешение на работу: Россия

Не готов к переезду, готов к командировкам

Желаемая должность и зарплата

NLP LLM Engineer

250 000 ₽ на руки

Специализации:

- Дата-сайентист
- Программист, разработчик
- Руководитель группы разработки

Занятость: полная занятость, частичная занятость, проектная работа, волонтерство, стажировка

График работы: полный день, сменный график, гибкий график, удаленная работа, вахтовый метод

Желательное время в пути до работы: не имеет значения

Опыт работы — 3 года 5 месяцев

Май 2024 —
настоящее время
1 год 3 месяца

Союзснаб

Москва, www.ssnab.ru/

Химическое производство, удобрения

- Агрохимия (продвижение, оптовая торговля)

Продукты питания

- Корма для животных (продвижение, оптовая торговля)
- Корма для животных (производство)

Сельское хозяйство

- Сельскохозяйственная продукция (продвижение, оптовая торговля)
- Земледелие, растениеводство, животноводство

AI engineer

СОЮЗСНАБ | AI Engineer

Ключевые достижения:

- Автоматизировал HR-процессы, сократив время создания поздравлений с 3+ часов до 30 секунд
- Повысил точность корпоративного поиска на 40% через внедрение LLM-ансамбля
- Масштабировал AI-решения на 4 отдела, обеспечив производительность 100+ операций/день

ПРОЕКТ 1: Корпоративная система интеллектуального поиска

Результат: Точность маршрутизации +40%, снижение ошибок выбора документов на 50%

Задача: Автоматизировать ответы сотрудников по внутренней документации, заменив менее точные методы поиска

Решение:

- Создал многоуровневый ансамбль локальных LLM (Llama 3: 7B + 70B модели)
- Внедрил семантический роутинг с голосующим механизмом по 6 подвопросам
- Развернул production-ready решение на GPU-сервере с миграцией на vLLM
- Дообучил embedding-модели на размеченной обратной связи с индексацией в Chroma

Масштаб: Успешно внедрено в 4 отделах компании с высокой масштабируемостью

ПРОЕКТ 2: Автоматизация HR-коммуникаций через Voice AI

Результат: Рост производительности HR с 5-10 до 100 поздравлений/день

Задача: Заменить ручной 3-часовой процесс создания персонализированных голосовых поздравлений

Решение:

- Разработал Telegram-бота с интеграцией Yandex SpeechKit для синтеза речи
- Реализовал автоматическое склонение имён через морфологический анализ
- Настроил прямую доставку в корпоративные системы коммуникации
- Исключил человеческие ошибки, повысив единообразие корпоративной коммуникации

Технический стек:

AI/ML: LLM Fine-tuning, RAG Systems, Semantic Search, Ensemble Methods, Voice AI
AI техники: Chain of Thought (CoT), Tree of Thought (ToT), Self-Reflection, Decomposition, CheckList, Structured Output (Pydantic)
Frameworks: LangChain, LangGraph, semantic_router, Hugging Face Transformers
Infrastructure: Docker, Ollama, vLLM, GPU Computing, NVIDIA GPU
Data: Vector Databases (Chroma, Pinecone), Embedding Models
APIs: Telegram API, Yandex SpeechKit, Python

Дополнительно: Реализованы множественные проекты с LLM-агентами и RAG-системами различной сложности. Детальное описание дополнительных проектов предоставляется по запросу.

Май 2023 — Май 2024
1 год 1 месяц

Академия астрологии Ирины Чайки

Москва, astrologchayka.ru/academy

Образовательные учреждения

- Тренинговые компании
- Обучение искусствам (рисование, пение, танцы, фото)

AI Engineer

ОНЛАЙН-ШКОЛА АСТРОЛОГИИ ИРИНЫ ЧАЙКИ | AI Engineer

Ключевые достижения:

- Увеличил конверсию продаж на 35% через персонализированный AI-консультант
- Автоматизировал 65% работы кураторов, сэкономив \$50K+ годовых расходов
- Сократил время получения ответов студентами с часов до 1 минуты

ПРОЕКТ 1: AI-консультант для автоматизации продаж

Результат: Рост конверсии на 35%, сокращение расходов на колл-центр на 42%

Задача: Повысить эффективность воронки продаж и снизить расходы на клиентский сервис

Решение:

- Разработал персонализированного чат-бота на базе GPT-4 для консультаций клиентов
- Создал AstrService для генерации натальных карт с автоматическими консультациями
- Внедрил цепочку из 9 специализированных LLM-вызовов для глубокой персонализации
- Автоматизировал создание PDF-отчетов с визуализацией и отправку клиентам

Impact: Значительный прирост выручки уже в первые 6 месяцев, высокая клиентская удовлетворенность

ПРОЕКТ 2: Автоматизация образовательного процесса

Результат: Сокращение времени кураторов на 65%, экономия ~30% годовых расходов на проверку

Задача: Автоматизировать проверку домашних заданий, которая занимала до 65% времени кураторов

Решение:

- Создал 4 специализированных чат-бота для проверки заданий по разным дисциплинам
- Реализовал гибридную систему поиска и проверки через Pinecone с морфологическим анализом
- Провел fine-tuning GPT-3.5 на 900 страницах контента и 1200 парах вопрос-ответ
- Применил продвинутые методы промпт-инженерии (CoT, ToT) для повышения точности

Масштаб: Стабильное качество проверок без человеческого фактора, повышение общей удовлетворенности

ПРОЕКТ 3: Корпоративная RAG-система для студентов

Результат: Время получения ответа < 1 минуты, существенное улучшение качества обучения

Краткое описание: Построил RAG-систему на тысячах страниц учебных материалов с семантическими коллекциями, иерархической картой знаний и поэтапным семантическим роутингом. Исключил "галлюцинации" LLM через точное соответствие исходному тексту.

Технический стек:

AI/ML: LLM Fine-tuning, RAG Systems, Conversational AI, Hybrid Search
AI техники: Chain of Thought (CoT), Tree of Thought (ToT), Self-Reflection, Decomposition
Frameworks: LangChain, LlamaIndex, Hugging Face Transformers
LLM: GPT-4, GPT-3.5 Fine-tuning, RAGAS Validation
Data: Pinecone, Chroma, BM-25, Vector Databases
Tools: PyMuPDF, Pymorphy2, JSON Structuring

Дополнительно: Реализованы системы аудита телефонных звонков, валидация RAG-ответов через RAGAS framework и другие AI-решения. Подробности по запросу.

Фриланс проекты

Москва

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения

Python Developer / ML Engineer (Freelance)

ПРОЕКТ: Telegram-бот "Учет расходов"

Результат: 1500+ активных пользователей, автоматизация учета личных финансов

Задача: Создать простого бота для автоматического учета и категоризации трат

Решение:

Разработал Telegram-бота на Python с базой данных SQLite

Обучил простую модель на логистической регрессии для классификации трат по категориям

Использовал ключевые слова и паттерны для автоматического определения типа расходов

Добавил визуализацию трат через простые графики

Создал систему уведомлений о превышении бюджета

Функциональность:

Автоматическое определение категории трат из текста

Экспорт статистики в Excel

Установка месячных лимитов по категориям

Простые графики расходов

Технический стек:

Backend: Python, aiogram, SQLite

ML: scikit-learn, логистическая регрессия, pandas

Visualization: matplotlib

Tools: Git, регулярные выражения

Образование

Высшее

2016

Пензенский государственный университет, Пенза

Факультет приборостроения, информационных технологий и систем, Автоматизация технологических процессов и производств (по отраслям)

Повышение квалификации, курсы

2025

Специализация "Data Engineering"

Coursera | AWS, "Data Engineering"

2024

GPT Week от Школы анализа данных (Яндекс)

Яндекс, AI Engineer

2024

Vector DataBase Master

Udemy, Data engineer

2023

Prompt Engineering for ChatGPT (Coursera)

Coursera (Университет Вандербилта), Prompt Engineer

2023

API ChatGPT Professional

Udemy, Prompt engineer

Электронные сертификаты

2025	Специализация "Data Engineering"
2024	GPT Week от Школы анализа данных

Навыки

Знание языков	Русский — Родной
Навыки	<div>SQLPythonLinuxGitDockerLLMLangChainREST API</div> <div>RAG SystemsFine-tuningPrompt EngineeringLangGraphHugging Face</div> <div>OllamavLLMGPT-4Text-to-SpeechMultimodal AIChroma</div> <div>PineconePydanticTelegram APIGPU ComputingBashSparkHive</div> <div>AirFlowKafkaHadoopOpenShift</div>

Опыт вождения

Имеется собственный автомобиль
Права категории B

Дополнительная информация

Обо мне	<p>AI Engineer с 3+ годами создания production LLM-решений и RAG-систем. Автоматизировал бизнес-процессы с impact \$50K+ экономии, повысил конверсию продаж на 35% и точность корпоративного поиска на 40%. Специализируюсь на conversational AI, voice технологиях и образовательных AI-платформах.</p> <p>Ссылка на мой git: https://github.com/Maskedxxx</p>
---------	--