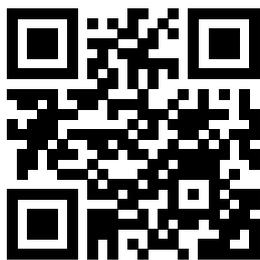


Ульяна Джумок — Data Scientist, Machine learning engineer



📍 Junior
📍 Россия, Новосибирск
💰 120 000 ₽ (руб/мес)
📅 Полная занятость
☁️ Удаленная работа
➔ Релокация
✉️ Контакт: geeklink.io/cv-124902



Навыки

Albumentations, AlexNet, Ansible, Auto-TS, AutoKeras, BiteTracker, Catboost, CircleCi, Classification, Computer vision, DeepSeek, Docker, EasyOCR, EfficientNet, FastAPI, Fine-tuning, GitLab, Google Text-to-Speech, GRU, Jenkins, Jiwer, Kubernetes, librosa, Llama, LLM, LoRA, LSTM, Matplotlib, Mistral, MobileNet, NLP, Numpy, Object Detection/Segmentation, OCR, OpenCV, PaddleOCR, Pandas, PEFT, PSPNet, PyAudio, Pytorch, ResNet, RNN, SAM, Scikit-learn, Seaborn, SpeechRecognition, SuryaOCR, TensorFlow/Keras, Tesseract, Timeseries Generator, torchvision, Transformers, U-Net, Unsloth, Uvicorn, VGG, ViT, WhisperAI, XGBoost, Yandex SpeechKit, YOLO.

Опыт работы

• Исследовательский проект

04.2025-04.2025 - Эксперименты с DeepSeek Janus Pro для мультимодальной обработки изображений, а также их генерации

- оптимизировала и квантизировала модель Janus Pro с использованием фреймворка Unsloth
 - реализовала мультимодальную обработку изображений с генерацией новых изображений по текстовому описанию
 - применяла Real-ESRGAN для увеличения размера и разрешения изображений, улучшая их визуальное качество
- Стек: Python, PyTorch, Deepseek Janus Pro

• Исследовательский проект

04.2025-04.2025 - Сравнение эффективности OCR и LLM моделей для задач распознавания русского текста

- провела исследование и протестировала различные подходы OCR-систем и LLM для обработки текстов на изображениях
 - оценила преимущества и ограничения каждого метода в применении к русскому печатному тексту, что позволило выбрать наиболее эффективные решения
- Стек: Python, OCR (Tesseract, EasyOCR, PaddleOCR, SuryaOCR), LLM API, Unsloth

• Прототип

01.2025-03.2025 - Автоматическая транскрипция видеоконференций и их суммаризация с LLM

- Дообучение моделей WhisperAI и DeepSeek R1 для повышения качества транскрипции русской речи и извлечения ключевых моментов из текста с выдачей результата в формате JSON.
- повысила точность транскрипции аудиоданных
 - реализовала извлечение ключевых моментов из транскрипции в структурированном виде
- Стек: Python, NLP, Hugging Face, Transformers, Unsloth

• «НМИЦ ТПМ» Минздрава России

01.2025 - 03.2025 - Проект «Разработка ИИ-системы, способной анализировать результаты биохимических

анализов крови по сканированным документам ФГБУ»

Создала REST API для загрузки изображений и получения текста с помощью различных OCR-систем
Стек: Python, FastAPI, Uvicorn, Flask, OCR, EasyOCR, Tesseract, PaddleOCR, SuryaOCR

- **Zinkworks, Ireland**

07.2020 - 08.2023 - DevOps Engineer

- обслуживала 20 Unix-серверов и поддерживала кодовую базу на Python
- автоматизировала обновление серверов с использованием Ansible, сократив влияние человеческого фактора
- настроила CI/CD пайплайн с Jenkins, снизив объём ручной работы на 80%
- выполняла ручное развертывание с использованием Docker и Kubernetes

- **Исследовательский проект**

12.2019-12.2019 - Классификатор вида цветов по изображению

Обучила модель для распознавания видов цветов
Стек: Python, Torch, Pillow, TorchVision

- **Исследовательский проект**

11.2019-11.2019 - Классификатор пород собак по изображению

- построила нейронную сеть для распознавания пород собак на основе моделей ResNet, AlexNet, VGG
 - достигла точности классификации на тестовых данных больше 90%
- Стек: Python, Torch, Pillow, TorchVision

- **БКС**

05.2016-03.2020 - Quality Assurance Engineer

- разработала тестовую архитектуру, стек технологий и рабочие процессы проекта
- реализовала и поддерживала функциональные тесты, автоматизировала тестирование с использованием SoapUI Pro
- проводила нагрузочное тестирование с применением Gatling и JMeter
- тестировала API различных типов: REST, GraphQL, WebSockets и gRPC

Образование

- **«Data Scientist, нейронные сети, машинное обучение и искусственный интеллект**

Университет искусственного интеллекта
03/2025

- **Master of Science in, Software design with Cloud computing**

Technological University of the Shannon, Ireland
2022

- **Программное обеспечение вычислительной техники и автоматизированных систем**

Сибирский государственный университет телекоммуникаций и информатики
2012

- **Machine Learning A-Z: AI, Python**

Udemy

2024

- **Statistics for Business Analytics and Data Science A-ZTM**

Udemy
2024

- **Cloud DevOps Engineer**

Udacity
2020

- **AI Programming with Python Nanodegree**

Udacity
2019

- **Автоматизация формирования коммерческого предложения на основе электрических схем**

EKF AI Challenge
2024

- **Green Cloud Computing: A review of the energy efficiency of deployment technologies**

Johnson Controls Tech Challenge
2022

Обо мне

Data Scientist с опытом в CV, OCR, NLP и немного LLM. Разрабатываю и дообучаю нейросети для обработки изображений, аудио и текста. Владею PyTorch, TensorFlow, Hugging Face, Unsloth и DevOps-инструментами. Умею сочетать анализ, исследование и инженерную реализацию для эффективных решений.