

# Kudrin Vladislav



Инженер машинного обучения



## Опыт работы

**Petrel AI (B2B продукты)** 📍 **Almaty**

Февраль 2022 - Апрель 2025 • 3 года 2 месяца



**Python Стажер** (4 месяца)

- ▶ Рефакторинг кода
- ▶ Исправление мелких багов
- ▶ Переписывание устаревшего кода



**ML Инженер** (2 года 10 месяцев)



**Примеры:**

- ▶ **Чат-бот** (RAG + Чат агент) (~98% точность правильных ответов) (llama + chromadb)
- ▶ **Нормализация текста** (Few shot обучение + RAG) (~88% правильного сопоставления и ~97% для ключ-значение) (qwen + chromadb)
- ▶ **Ревьюер коммитов** (API интеграция + Код агент) (~0.83 F1) (claude)
- ▶ **AI репетитор** (TTS + STT + Парсинг изображений + Поисковый агент) (нет метрик) (claude + llava + whisper + minimax audio)
- ▶ **Детекция спама** (дообучение BERT) (~0.99 F1 для NSFW, ~0.82 F1 для рекламы) (debert)
- ▶ **HR Агент** (API интеграция + Парсинг документов + Скоринг) (~0.71 F1 правильной классификации, в основном из-за человеческого фактора) (claude + llava)

# Образование

---

## Уральский федеральный университет (2021 - 2025)

- ▶ Степень бакалавра в области компьютерных наук
- ▶ Дополнительное образование в области ИИ

# Языки

---

- ▶ Русский (родной)
- ▶ Английский (свободно)

# Технические навыки

---

- ▶ Создание python микросервисов
- ▶ Создание AI агентов, RAG пайплайнов
- ▶ Работа с моделями генерации текста, text2text, классификации
- ▶ Работа с базами данных
- ▶ Дообучение моделей
- ▶ Prompt инжиниринг
- ▶ Развертывание, работа с облаком
- ▶ Интеграция с другими сервисами

# Мягкие навыки

---

- ▶ Ответственность
- ▶ Преданность делу
- ▶ Адаптивность
- ▶ Быстрое обучение

## Языки программирования

- ▶ Python (основной)
- ▶ Javascript (немного)
- ▶ C# (немного)
- ▶ C++ (немного)

## Базы данных

- ▶ PostgreSQL
- ▶ MongoDB (для документоориентированных)
- ▶ ChromaDB (продакшн или тестирование)
- ▶ FAISS (продакшн)

## Агенты

- ▶ Langchain (организация пайплайнов)
- ▶ Langgraph (создание агентов)
- ▶ Langfuse (мониторинг агентов)
- ▶ PydanticAI (создание агентов)
- ▶ SmolAgents (создание легких агентов)
- ▶ LlamaIndex (интеграция RAG)

## Глубокое обучение

- ▶ PyTorch (дообучение)
- ▶ Unsloth (дообучение)
- ▶ PEFT (дообучение)
- ▶ Transformers (тестирование)
- ▶ Sentence transformers (эмбединги)

## Backend и DevOps

- ▶ FastAPI (REST API)
- ▶ Docker (образы, compose, контейнеры)
- ▶ Kubernetes (практика в minikube)
- ▶ Git (github/gitlab)
- ▶ CI/CD (github actions)

## MLOps

- ▶ TGI (продакшн)
- ▶ vLLM (продакшн)
- ▶ exllamav2 (продакшн)
- ▶ llama.cpp (тестирование)
- ▶ ollama (тестирование)

## Модели

- ▶ Все opensource модели генерации текста (DeepSeek, qwen, mistral, llama, gemma...)
- ▶ Закрытые модели генерации текста (GPT, Claude, Grok, Gemini)
- ▶ text2text и классификация (DistilBERT, RoBERT, DeBERT, BART, T5, opus)

## Облачные сервисы

- ▶ Google cloud (объектное хранилище, serverless контейнеры, балансировщики нагрузки, sql кластеры, firebase)
- ▶ Yandex cloud (для России)
- ▶ Openrouter (для отказоустойчивости и удобства использования)
- ▶ DeepInfra, NovitaAI, Groq (прямое более дешевое использование)

## Мультимодальность

- ▶ TTS (minimax audio API и kokoro)
- ▶ STT (whisper)
- ▶ Изображение в текст (llava, gemma, mistral small в основном для парсинга)
- ▶ Видео, аудио (gemini flash)

## IDE

- ▶ Cursor (сейчас использую claude 4)
- ▶ Windsurf (иногда)
- ▶ Jupyter (для дообучения или крайних случаев)

## Frontend

- ▶ Streamlit
- ▶ React (немного)

## ОС

- ▶ Linux (для работы и личного использования)

## Также

- ▶ Asyncio (python)

## Дополнительно

---

- ▶ Создал AI визуальную новеллу в браузере [redcamptale.web.app](https://redcamptale.web.app)
- ▶ Занял 1-е место в хакатоне LLM coding challenge от сбербанка (сельскохозяйственный кейс)
- ▶ Вношу вклад в open source проекты (ollama, pydanticAI)
- ▶ Читаю статьи на arxiv (недавно была codeact)
- ▶ Слежу за новостями в X\* (заблокирован в России), Youtube каналах (fireship, bycloud) и моделями hugging face (трендовые)
- ▶ Использую Claude 4 как помощника для кодирования.

## Ссылки

---

 **Github** (некоторые пет-проекты)


 **Kaggle** (участие в соревнованиях, всегда топ 10%)


 **Hugging Face** (есть слияния и GGUF конвертации)

 **LeetCode** (python, C#, SQL)

## Контакты

---

 **Telegram**  
@vladlen32230

 **Gmail**  
vladlen32230@gmail.com