Dmitrii Kuzmin

NLP/ML Engineer · NLP Researcher

✉ 1kkiren@mail.ru | ⌂ 1kkiren.ru | ⬛ 1kkiRen

Summary

NLP Engineer with over a year of both research and industry experience. A good team worker, passionate about optimizing processes and finding effective solutions for AI development. Skilled in Python, PyTorch, and model fine-tuning. Experienced in working with open-source models and interested in adapting and optimizing LLMs.

Skills

NLP	PyTorch, Transformers, Tokenizers, LangChain, numpy, pandas
DevOps	Git, Docker
Back-end	MongoDB, Telegram API
Languages	English, Russian (native)
Soft skills	Flexibility, Responsibility, Enthusiasm

Work Experience

Mohamed bin Zayed University of Artificial Intelligence   *Abu Dhabi, United Arab Emirates*

RESEARCH INTERN   *June. 2025 - present*

- Developing the alternative tokenization methods.
- Writing the paper.

DeepPavlov   *Moscow, Russia*

MIDDLE NLP ENGINEER   *May. 2025 - present*

- R&D
- Conducting LLM testing.
- Working with GPUs of Chinese manufacturers.

Center for Applied AI, Skolkovo   *Moscow, Russia*

MIDDLE NLP ENGINEER   *Feb. 2025 - May 2025*

- Fine-tuned the Qwen2.5-VL model.
- Pipeline development.
- Selection of prompts for generating remarks for different types of requirements.

Higher School of Economics   *Moscow, Russia*

NLP RESEARCHER   *Jun. 2024 - May 2025*

- Fine-tuned Llama3-8B-Instruct model.
- Developed Russian BPE tokenizer.
- Developed the solutions to manipulate the existing tokenizer, e.g. delete existing tokens.
- Created grammar benchmark.

Moscow Aviation Institute   *Moscow, Russia*

ML ENGINEER/BACKEND ENGINEER   *Jul. 2023 - Oct. 2023*

- Developed sentence theme classification model.
- Optimized database queries.
- Worked with Telegram API.

Innopolis Universty   *Innopolis, Russia*

NLP ENGINEER   *Jun. 2023 - Jul. 2023*

- Developed Deep Learning model for sentiment analysis of YouTube comments.
- Optimized and fine-tuned BERT model.

Papers

Rethinking Tokenization: Improving Language Model Performance by Modifying the Input Module at Inference   *EACL 2026*

RESEARCHER & WRITER   *Jun. 2025 - present*

- The paper demonstrates the influence of the alternative tokenizations of the same text on the quality of model inference.
- **Under review.**

TokenSubstitution: Cost-efficient method for LLM adaptation for the Russian language *ACL 2026*

RESEARCHER & WRITER *Feb. 2025 - present*

- The paper suggests cost-effective adaptation approach for improving the performance of LLM in target language.
- **In progress.**

A Multi-Aspect Evaluation of Tokenizer Adaptation Methods For Lagre Language Models on Russian *AI Journey 2025*

RESEARCHER & WRITER *Jun. 2024 - Jun. 2025*

- The paper demonstrates that tokenizer adaptation is a cost-effective and efficient approach for improving LLMs, considering text quality and token efficiency in multiple benchmark datasets.
- **Accepted to the conference.**

Projects

Tokenizer Changer

LIBRARY *Jul. 2024 - present*

- The open-source library for changing existing 🤗 tokenizers.
- `https://pypi.org/project/TokenizerChanger/`
- `https://github.com/1kkiRen/Tokenizer-Changer`

Embeddings Division

LIBRARY *Jan. 2025 - present*

- An open source library for separating Embedding layers in large language models.
- `https://pypi.org/project/embdiv`
- `https://github.com/1kkiRen/Embeddings-Division`

CRUD Calendar LLM Chatbot

FREELANCE *Feb. 2025*

- Telegram-chatbot that works as a calendar manager and can provide an up-to-date news summary.

Education

Innopolis University *Innopolis, Russia*

B.S. IN DATA ANALYSIS AND ARTIFICIAL INTELLIGENCE *2022 - 2026*

- Core Courses: Software Systems Analysis and Design, Human AI Interaction Design, Mathematical Analysis

Extracurricular Activity

Innopolis University *Innopolis, Russia*

TUTOR *Sep. 2023 - Jan. 2024*

- Helped 1st Bachelor's students adapt to the University.
- Organized extracurricular events.