

Shanghai, China or
(currently) Moscow, Russia
+8613621053870

Nikolai Zakharov

n.d.zakharov@outlook.com
<https://www.linkedin.com/in/nikolai-zakharov-285841a7/>

Professional Experience

Machine Learning Engineer, **VK** **Jan 2025 – Present**
VkVideo

- Built a TV series detection, normalization, and metadata extraction pipeline using **regex bootstrapping** and **LLM-generated data**. Improved the **similar video recommendation system** on the **VK Video** platform, leading to a consistent **4.5–6.3% increase** in view counts and total watch time across short (30s/60s) and full-length formats.
- Identified and fixed scroll session bugs (6% cases), enhancing session tracking and doc2doc relevance.
- Deployed geo-ban and music filters, ensuring compliance and improving anchor-based music recommendations.

Machine Learning Engineer, **NIO** **Jul 2023 – Jan 2025**
AD

- Developed a **RAG-based assistant** for driving log analysis, accelerating root-cause investigations.
- Built **LLM-based classifiers** to assess idea quality, reducing reliance on SMEs. Provided **automatic feedback**, increasing acceptance rate by **20%**.
- Developed a C++ module for real-time autonomous driving event detection, deployed in live vehicles to enable early identification of edge-case scenarios like slow car overtakes.
- Reduced **Apollo preprocessing time by 50%** with a custom Python converter.

Machine Learning Engineer, **Tinkoff (T-Bank)** **Sep 2021 – Jan 2023**
Voice Kit

- Deployed a Russian Voice Conversion model, achieving **95% similarity** and improving voice assistant quality.
- Reduced Real-Time Factor by 88% for a production HifiGAN vocoder, **enhancing response times**.

Machine Learning Engineer, **Huawei Technologies** **Jul 2019 – Sep 2021**
CBG

- Built a multilingual trie-based NLU engine and automatic model update service with **80%+ request coverage**, cutting inference cost by 20%.
- Built a voice cloning model for cloud and on-device environments with fine-tuning support.

Research Intern, Noah Ark **Huawei Technologies** **2018 – 2019**
Lab

- Worked on AutoML (DARTS) and GAN-based solutions for CV and face recognition datasets.

Technical Skills and Languages

- Languages: Python (PyTorch, TensorFlow, HuggingFace), C++, Go, Java.
- NLP/GenAI: GPT fine-tuning, RAG, semantic search, tokenization, voice modeling.
- Infra: Docker, Kubernetes, KServe, Helm, Prometheus, Grafana, distributed inference.
- Spoken Languages: English (Fluent), Chinese (Fluent), Russian (Native), Turkish (Basic).

Education

Tsinghua University **Beijing, China** **Sep 2016 – Jul 2019**

- **Master of Science in Computer Science, Machine Learning.**
- **Thesis:** Diversified Image Captioning with Deep Learning (written in Chinese).
- Awarded a full Chinese Government Scholarship for academic excellence.

Publications and Open-Source Contributions

- **Paper:** Towards Controllable Image Descriptions with Semi-Supervised VAE (Zakharov, Nikolai). Published in *Journal of Visual Communication and Image Representation*, 63 (2019): 102574. <https://doi.org/10.1016/j.jvcir.2019.102574>
- **GitHub:** Open-source contributions available at <https://github.com/yiyang92>.