

# Dmitry Barsukov

---

CONTACT INFORMATION	Telegram: <a href="#">@riZZZhik</a> (preferred) GitHub: <a href="#">github.com/riZZZhik</a> LinkedIn: <a href="#">linkedin.com/in/riZZZhik</a> Email: <a href="#">riZZZhik@gmail.com</a>	Residence: Moscow, Russia Open to remote and hybrid full-time jobs
PROFESSIONAL SUMMARY	Experienced Machine Learning Engineer with 6 years in Data Science focusing on Text-to-Speech, Computer Vision, and Deep Learning optimization.	
PROGRAMMING SKILLS	<b>Languages:</b> Python (advanced), Go (advanced), Tritonlang (average), C/C++ (average), CUDA C++ (average) <b>Deep Learning Frameworks:</b> PyTorch, TensorFlow, Keras, scikit-learn <b>Deep Learning Architectures:</b> Transformer, Diffusion, GAN, YOLO, Speech SOTA <b>Optimization frameworks:</b> TensorRT, torchcompile, Tritonlang, OpenVINO, ONNX Runtime, LiteRT (f.k.a. TensorFlow Lite) <b>Technical skills:</b> Triton Server, OpenCV, ffmpeg, Torchaudio, WandB, ClearML, Docker, Kubernetes, Git, CI/CD, Observability, Prometheus, Grafana, Linux	
LANGUAGES	Russian (native); English (advanced)	
EMPLOYMENT AND EXPERIENCE	<b>Victory Group</b> Machine Learning Team Lead	April 2025 - Present
	<p>Team lead for a team of 4 data scientists and machine learning engineers, responsible for the development of a multimodal (text, audio, classic) AI systems for real-time text and audio processing.</p> <p><b>Responsibilities:</b></p> <ul style="list-style-type: none"><li>– Leading the team in the research, development and deployment of a multimodal AI system.</li><li>– Mentoring and guiding team members in their professional growth.</li><li>– Collaborating with cross-functional teams to ensure successful project delivery.</li></ul> <p><b>Achievements:</b></p> <ul style="list-style-type: none"><li>– Hired and mentored a team of 4 DS and MLE.</li><li>– Built from scratch all the processes and infrastructure for the team, including:<ul style="list-style-type: none"><li>– Model development and deployment processes</li><li>– CI/CD pipelines for model deployment</li><li>– Version control and artifact management</li><li>– Monitoring, logging and alerting</li></ul></li><li>– Under my leadership, the team has successfully developed and deployed:<ul style="list-style-type: none"><li>– A multimodal AI system combining ASR, TTS, and LLM components for real-time phone conversations that achieve human-level performance.</li><li>– A comprehensive AI solution for automated phone conversation quality assessment and analysis.</li></ul></li></ul>	

**MTS AI**  
Senior Machine Learning Engineer

June 2022 - March 2025

Development of a Text-to-Speech, Speech-to-Text, and ASR services that outperform leading competitors in the Russian language.

**Responsibilities:**

- Model deployment using Triton Server, Docker, Kubernetes, Python, and Golang.
- Model optimization for performance and resource efficiency:
  - Model architecture changes
  - TensorRT, tritonlang, torchcompile and OpenVINO
  - Model warmup, quantization, sparsity and pruning
- Research, develop, train and fine-tune new model architectures.

**Achievements:**

- **0.12** p95 latency and **90** RPS for diffusion model on a single **2g.20Gb** A100 instance.
- Established and automated a version-controlled model deployment process using CI/CD, WandB / ClearML, and Artifactory.
- The development process was established following best practices (e.g., CI/CD, code review, documentation, unit testing, changelog, semantic versioning, etc.).
- Created an automated quality and performance testing in a production-like Kubernetes environment.
- Implemented comprehensive observability with monitoring, logging, and alerting.

**Technologies:**

Python, Golang, PyTorch, tritonlang, CUDA C++, torchcompile, TensorRT, WandB / ClearML, Triton Server, Observability, Docker + Kubernetes, Git + CI/CD

**SIRIN**  
Senior Machine Learning Developer

March 2021 - January 2022

Designed and implemented machine learning service for the automatic opening of car barriers using computer vision.

Achieved **99%** accuracy in recognizing Russian license plates and **90%** accuracy for all other license plates, maintaining a latency of **0.5** on a **4-core** CPU.

**Technologies:**

Python, PyTorch, OpenCV, Docker + Kubernetes, OpenVINO + Triton Server, Observability (Grafana, Kibana, Prometheus), Git + CI/CD

**ITMO University**  
Machine Learning Developer

January 2020 - December 2020

Designed and implemented service for building facade segmentation, managing everything from data collection and preprocessing to model training, evaluation, and deployment.

**Technologies:**

Python, TensorFlow + Keras, OpenCV, Docker, Git

**SPIIRAS**

August 2018 - October 2020

Middle Machine Learning developer

Designed and implemented facial recognition service, managing everything from data collection/generation and preprocessing to model training, evaluation, and deployment.

**Technologies:**

Python, TensorFlow + Keras, RealSense DepthCamera, OpenCV, Docker, Git

## EDUCATION

**Higher School of Economics**

Moscow, Russia (Remote)

B.S., Applied Mathematics and Information Science.

September 2023 - Present

*GPA: 3.7/4.0*