

MARAT GASANOV

ML Engineer

✉ maratgasanov6@gmail.com 📞 +7 989 881-13-36 ✈ @M4PAT1K
🌐 maratik.fun 🏠 github.com/gassannov 🔗 linkedin.com/in/marat-gasanov-a82878308

PROFILE

ML engineer building text and image models from data preparation and training to production. Focused on generative AI, domain-specific translation, LLM agents and inference performance.

EXPERIENCE

Ozon | ML Engineer | 11/2024 - Present

- Owned inference optimization from requirements to load testing. Introduced asynchronous batching in Triton and integrated vLLM; increased throughput from **20 to 200 RPS (10x)** without additional infrastructure scaling.
- Built a Kazakh translator for product descriptions and UGC: prepared domain data, trained models and evaluated difficult cases. Domain MQM score: **0.87 vs 0.77** for Yandex Translate.
- Built and launched reference-based background generation with VLM and image-edit models; fine-tuned for product-detail preservation. Reached **80%** on the internal aesthetics metric; saved-result share reached **70%** in the product.
- Improved a dialog-topic agent with the team: curated few-shot examples and tested prompts, long context, reasoning and tools. Quality increased by **20%**; the agent became **1.5x faster**.

Art. Lebedev Studio | ML Engineer | 11/2023 - 11/2024

- Developed ML components for Ironov 2.0/3.0. Trained and integrated BERT to extract visual concepts from client briefs; owned data preparation, training and pipeline integration.
- Sourced datasets and trained **30+ visual styles** with LoRA on SDXL, Kandinsky and FLUX.1-dev; integrated them into logo generation. Service scale at the 3.0 release: **3M+ generated logos**. Presented updates at the Studio's Summer Days 2024.

SberBank | Data Analyst | 08/2023 - 11/2023

- Built transaction features and anomaly scoring for audit with Python, SQL, ClickHouse, CatBoost and Airflow. Processed **~20M transactions**; flagged **3,000 operations** for manual review. The solution helped uncover violations in **3 departments**.
- Built a rental-price validation prototype with LightGBM and PostgreSQL: checked **~75,000 contracts** in a few hours and identified **~900 properties** with substantial market deviations.
- Built an OCR document-validation prototype with OpenCV, FastAPI and Docker. At **~12,000 documents/month**, **~85%** passed initial checks automatically; the manual queue was **~3x smaller**.

redrocket.ai | ML Engineer | 01/2023 - 08/2023

- Developed and deployed an image-generation model; integrated RabbitMQ and implemented load balancing. Stack: Python, Stable Diffusion, RabbitMQ, VastAPI.

EDUCATION

Bauman Moscow State Technical University | 2020 - 2024

VK Education at Bauman Moscow State Technical University | 01/2022 - 06/2023

TECHNICAL SKILLS

Python, SQL, BERT, LLMs, VLMs, LoRA, SDXL, Kandinsky, FLUX.1-dev, ComfyUI, Triton Inference Server, vLLM, CatBoost, LightGBM, Airflow, OpenCV, FastAPI, Docker, PostgreSQL, ClickHouse