ROMAN GEVORGYAN

r.gevorgyan@g.nsu.ru | +7-905-083-7168 | t.me/rromanofficial | GitHub

High-intensity ML Engineer and Yandex developer specializing in Generative AI (Stable Diffusion, Flux, LLMs) and automated content pipelines. Expert in building end-to-end production systems for AI-driven platforms, including content filtering and distributed generation. Proven ability to deliver rapid results in high-pressure environments, specializing in FastAPI, Django, and Telegram Bot orchestration for rapid product prototyping.

EXPERIENCE

Yandex - *ML Engineer (Generative AI & AliceAI)* — Mar 2025 - Present

- Architected full production pipelines for high-quality ad banner generation using Chain-of-Thought (CoT) and iterative LLM distillation, improving relevance and semantic diversity.
- Built automated content filtering and deduplication systems using classification models to ensure visual consistency and relevance for commercial platforms.
- Developed end-to-end reward models for "Neuro Alice" to perform structured verification (factuality, risks, sources), enhancing conversational AI reliability.
- Optimized distributed GPU inference and MapReduce workflows (YQL over YT, Nirvana) to scale content processing and generation speed.

SHAD / Sirius University - *AI Researcher (Generative Systems)* — 2023 - 2025

- Optimized Stable Diffusion workflows using DPM-Solver, LoRA, and 4-bit quantization, achieving high-fidelity generation with minimal sampling steps (10-20).
- Designed and trained diffusion models for panoramic view generation and neural game simulation utilizing GameGAN and World Models for interactive environments.
- Researched SOTA architectures for real-time video simulation, directly applicable to automated AI-influencer content workflows and virtual environment platforms.

PROJECTS & ACHIEVEMENTS

1st Place - CV & Digital Services Hackathon (NTO)

- Built a full-stack platform for visual location search and automated route generation, demonstrating rapid prototyping of web services under strict deadlines.

Neural Content Generation Framework

- Developed automated content workflows utilizing Flux, Stable Diffusion, and custom LoRAs for consistent character generation across influencer platforms.
- Integrated backend services with Telegram Bot API and FastAPI to create a seamless user interface for rapid content generation and management.

SKILLS

Generative AI: Stable Diffusion, Flux, LoRA, ControlNet, IP-Adapter, Flow Matching, LangChain, Transformers, HuggingFace, Reward Models, YandexGPT API

Backend & Bot Dev: Python, FastAPI, Django, Flask, SQL (PostgreSQL, YQL), Redis, Docker, MLFlow, MapReduce, Nirvana, Telegram Bot API

Core ML: PyTorch, TensorFlow, CatBoost, XGBoost, Scikit-learn, Pandas, NumPy, Statistics

EDUCATION

Yandex School of Data Analysis (SHAD) - *Data Science Specialization* — 2023 - 2025

Novosibirsk State University - *BS/MS Mechanics and Mathematics* — 2021 - 2026