Igor Pavlov

+7-925-369-20-48 ivladipavlov@gmail.com boomb0om

Experienced Machine Learning and Python Engineer with a strong background in building large-scale data processing systems, web scrapers, and AI-driven solutions. Proven expertise in developing multimodal data pipelines, optimizing model inference, and deploying production-ready AI models. Passionate about system design, data governance, and enhancing machine learning workflows. Always eager to tackle complex challenges and drive innovation in AI and software engineering.

Work experience

X-Labs AI **August 2024 – April 2025**

Lead Python Engineer *Moscow*

- Led the development of a platform for providing web scrapers via API—an analogue of the Oxylabs/BrightData web-scraping API. The platform serves three B2B clients with a total order volume exceeding 1 million RUB;
- Led the development of a system for processing and storing multimodal data based on S3, Airflow, and FastAPI. Implemented data governance practices: developed a data catalog for ML teams and wrote data tests. This improved data quality and simplified the work of ML teams;
- Developed a pipeline for collecting and preprocessing audio data from more than 20 sources. The collected dataset was used to train a song generation model with vocals and a beat generation model. These models became the core feature of several company products;
- Led a team of three engineers, represented the company at conferences and meetups, and pitched projects.

Sber AI **June 2021 – August 2024**

Machine Learning Engineer *Moscow*

- Developed ETL pipelines for collecting and processing multimodal data for training text2image, text2video, and other models. Collected a dataset used to train the Kandinsky and Kandinsky Video models;
- Developed and published a framework for processing and filtering image and video data. This framework is used by various teams at Sber. Open-source version, article;
- Developed web scrapers for collecting textual, video, and image data;
- Designed and developed inference for the Kandinsky and Kandinsky Video models. Collaborated with the backend team and deployed these models to production;
- Optimized model inference in PyTorch, achieving a 15% speedup for a diffusion model;
- Developed a benchmark for reproducible evaluation of text2image models. Open-source version;
- Trained a watermark detection model for images with an accuracy of 93.44%, compared to 77.86% for the SOTA open-source model. The model was used for dataset filtering. Code;
- Developed a model for removing watermarks from images. The model is used for processing images and videos in datasets.

Sber AI **August 2020 – June 2021**

Intern Machine Learning Engineer *Москва*

- Researched various approaches to style transfer between images;
- Trained an image quality enhancement model based on Real-ESRGAN. The model was later used in the image generation pipeline and other internal projects. GitHub repository gained over 500 stars.

Education

ITMO, AI Talent Hub **September 2024 – Present**

Master's Program in "Artificial Intelligence" *Moscow*

Bauman Moscow State Technical University, Department IU9 **September 2020 – July 2024**

Bachelor's in Theoretical Informatics and Computer Technologies *Moscow*

Karpov Courses, System Design **October 2023 – December 2023**

System Design *Moscow*

Completed a course on designing high-load systems. Certificate.

Techincal skills

Programming languanges: Python, SQL
Databases and message brokers: PostgreSQL, MongoDB, Redis, Kafka, RabbitMQ, NATS
ML: PyTorch, Transformers, Diffusers, CatBoost, sklearn
Python: asyncio, fastapi, multiprocessing, threading, boto3, fsspec, fastapi, pytest, mypy
Other: Linux, Docker, Airflow, S3, Bash, Git
Languages: English B2, Russian Native