Leonid Matyushin

leonid.sh

Email : matyushinleonid@gmail.com

Telegram : @matyushinleonid

Software Engineer — backend, distributed systems, cloud and ML infrastructure. Experience building production-grade software at Nebius and NtechLab.

Education

- **Skoltech / Higher School of Economics** — Moscow, Russia
  MSc, CDISE / Faculty of Computer Science — *Sep 2018 – Jun 2020*
- **Higher School of Economics** — Moscow, Russia
  BSc, Faculty of Mathematics — *Sep 2014 – Jul 2018*
- **AESC, Lomonosov Moscow State University** — Moscow, Russia
  Final Year of Secondary School — *Sep 2013 – Jun 2014*

Experience (Selected Projects)

- **Nebius** — Amsterdam, Netherlands
  Software Engineer — *Aug 2023 – Nov 2025*
  - **HWaaS: Disk Replacement Automation**: Designed and implemented infrastructure that fully automated failed-disk replacement, in particular the creation of Jira tickets for data-center engineers. I built the solution as part of a Kubernetes operator and added a mechanism that disables automatic replacement when monitoring behaves anomalously. The system has performed 1,000+ disk replacements across 4+ regions for YDB and an in-house S3-compatible object storage. Extended the operator's gRPC API, performed CRD migrations, agreed failure scopes and regional rollout order with SREs and managers, and set up monitoring and alerting. **Go, Kubebuilder**
  - **HWaaS: Automated Deployment**: Designed and implemented an automated Kubernetes build and deployment process for the HWaaS service, which ships as a set of tightly coupled Kubernetes resources. Adopted GitOps delivery so releases became reversible, faster, and transparent to other developers. **Kubernetes, TeamCity, Timoni, Helm, Argo CD, Argo Workflows**
  - **Dubformer: Media Processing Backend**: Designed and implemented machine translation and text-to-speech microservices and integrated them into the automatic video translation pipeline. Developed REST APIs for individual services and for the pipeline as a whole. **Python, Airflow, FastAPI, MongoDB, AWS SQS, AWS S3**
  - **Dubformer: Infrastructure Engineering**: Built monitoring and alerting that exposed backend behavior and load through dashboards and enabled rapid incident response through messenger notifications. Implemented CI/CD pipelines and reduced the total build and test time from 60 to 10 minutes. **Docker, GitHub Actions, Grafana, Prometheus**
- **NtechLab** — Moscow, Russia
  Software Engineer — *Nov 2020 – Aug 2023*
  - **Speed and Memory Benchmarking Service**: Designed and implemented an internal service for benchmarking inference speed and memory consumption of computer vision models developed in the R&D department. Exposed a REST API and built CLI and SDK interfaces on top of it. Proactively collected feedback and drove adoption across multiple teams, including object detection, face recognition, and license plate recognition, significantly reducing research iteration time. **Python, Celery, PostgreSQL, RabbitMQ**
  - **Data Labeling Pipelines**: Developed Airflow pipelines that automated data-labeling workflows for image-classification tasks in the R&D department. **Python, Airflow, Yandex Toloka**
  - **NSFW Image Classification**: Developed an NSFW content-detection model that outperformed popular open-source alternatives on public and internally created benchmarks, then deployed it to customer infrastructure. **Python, PyTorch, PyTorch Lightning, TorchServe, Albumentations**
  - **Falling People Detection**: Developed a falling-people detection model using a domain-adapted neural network on top of an open-source skeletal-pose model. Achieved a high true-positive rate at an extremely low false-positive rate on benchmarks I collected myself, and owned the lifecycle from data collection and labeling through product integration. **Python, PyTorch, TensorRT, OpenVINO**
  - **Acceleration of the Facial Recognition Model**: Accelerated the GPU face-recognition model by 10% without accuracy loss by using neural-architecture-search techniques to identify a better backbone. **Python, PyTorch**

- **Laboratory of Methods for Big Data Analysis** — Moscow, Russia
  Research Intern — *Jan 2020 – Nov 2020*
- **Deeplight Ventures** — Moscow, Russia
  Software Engineer — *Feb 2019 – Feb 2020*
- **Sberbank** — Moscow, Russia
  Data Science Intern — *Aug 2017 – Sep 2017*

Personal Projects

- **k8s.leonid.sh**: A personal Kubernetes platform on Yandex Cloud. Cloud resources — cluster, network, managed PostgreSQL, registry, DNS, secret storage — are declared as code in Terraform and bootstrapped with Ansible; every workload on top is delivered by Argo CD from Git. The platform layer carries Envoy Gateway with cert-manager and ExternalDNS for ingress and certificates, External Secrets for credentials, and self-hosted Prometheus, Grafana, and Loki for monitoring, alerting, and logs. It hosts the rest of the projects here. **Kubernetes, Terraform, Ansible, Argo CD, Envoy, Prometheus**
- **Sein zum Tode**: A Telegram bot intended to help users confront the prospect of death as something that will happen and, moreover, **is already happening** to them. The user is asked a series of questions, after which an LLM estimates roughly how much of their life is left. The value of it is the daily shrinking number of remaining days, which the bot keeps sending. Each conversation runs as a Temporal workflow, PostgreSQL is the main database, and Redis keeps the sensitive answers only temporarily, for privacy. **Python, aiogram, Temporal, OpenAI API, PostgreSQL, Redis**

Other Achievements

- **Rosneft Seismic Challenge**: International computer vision competition in seismic image segmentation. Achieved 2nd place out of 98 participating teams (boosters.pro, Oct 2019 – Dec 2019).
- **System Design Course**: Completed an advanced course on designing distributed, high-load systems and graduated with honors (karpov.courses, Jan 2023 – Feb 2023).
- **Strength Training and Physical Transformation**: Achieved a long-term body recomposition by losing over 28 kg of body weight while building significant muscle mass, demonstrating discipline, consistency, and goal-oriented training (Apr 2024 – present).

Technical Skills

- Go, Python
- AWS, Yandex Cloud, Hetzner Cloud, Nebius Cloud
- Kubernetes, Docker, Docker Compose, Helm, Timoni, Argo CD, Argo Workflows
- Git, GitHub Actions, GitLab CI, TeamCity
- Airflow, FastAPI, Flask, Temporal
- PostgreSQL, MongoDB, YDB, Redis, RabbitMQ, AWS SQS
- Prometheus, Grafana, Loki
- PyTorch, TensorFlow, PyTorch Lightning, TorchServe, OpenVINO, TensorRT, Pandas, NumPy, scikit-learn, CatBoost, XGBoost
- English, Russian