Леонид Матюшин

leonid.sh

Email : matyushinleonid@gmail.com

Telegram : @matyushinleonid

Software Engineer — бэкенд, распределённые системы, облачная и ML-инфраструктура. Опыт промышленной разработки в Nebius и NtechLab.

Образование

- **Сколтех / Высшая школа экономики** — Москва, Россия
  Магистратура, CDISE / факультет компьютерных наук — *сен 2018 – июн 2020*
- **Высшая школа экономики** — Москва, Россия
  Бакалавриат, факультет математики — *сен 2014 – июл 2018*
- **СУНЦ МГУ имени М. В. Ломоносова** — Москва, Россия
  11-й класс школы — *сен 2013 – июн 2014*

Опыт работы

- **Nebius** — Амстердам, Нидерланды
  Software Engineer — *авг 2023 – ноя 2025*
  - **HWaaS: автоматизация замены дисков**: Спроектировал и реализовал инфраструктуру, полностью автоматизировавшую замену неисправных дисков (в частности, создание Jira-тикетов для инженеров дата-центров). Решение я построил как часть Kubernetes-оператора; также был добавлен механизм, отключающий автоматическую замену при аномальном поведении мониторинга. Система выполнила более 1 000 замен дисков в 4+ регионах для YDB и внутреннего S3-совместимого object storage. Расширил gRPC API оператора, провёл миграции CRD, согласовал сценарии отказов и очерёдность регионального запуска с SRE и менеджерами, настроил мониторинг и алертинг. **Go, Kubebuilder**
  - **HWaaS: автоматизированный деплой**: Спроектировал и реализовал автоматическую сборку и разворачивание сервиса HWaaS, который поставляется как набор тесно связанных Kubernetes-ресурсов. Внедрил GitOps-подход, благодаря которому релизы стали обратимыми, быстрее и прозрачнее для остальных разработчиков. **Kubernetes, TeamCity, Timoni, Helm, Argo CD, Argo Workflows**
  - **Dubformer: бэкенд обработки медиа**: Спроектировал и реализовал микросервисы машинного перевода и синтеза речи и интегрировал их в автоматический пайплайн перевода видео. Разработал REST API как отдельных сервисов, так и пайплайна целиком. **Python, Airflow, FastAPI, MongoDB, AWS SQS, AWS S3**
  - **Dubformer: инфраструктура**: Построил мониторинг и алертинг, отображающий поведение и нагрузку бэкенда на дашбордах и ускоряющий реакцию на инциденты через уведомления в мессенджере. Реализовал CI/CD и сократил суммарное время сборки и тестов с 60 до 10 минут. **Docker, GitHub Actions, Grafana, Prometheus**
- **NtechLab** — Москва, Россия
  Software Engineer — *ноя 2020 – авг 2023*
  - **Сервис измерения скорости и памяти**: Спроектировал и реализовал внутренний сервис для измерения скорости инференса и потребления памяти моделями компьютерного зрения, которые разрабатывались в R&D-подразделении. Предоставил REST API и построил поверх него CLI и SDK. Проактивно собирал обратную связь и добился внедрения в нескольких командах — детекции объектов, распознавания лиц и автомобильных номеров, — значительно сократив цикл исследовательских экспериментов. **Python, Celery, PostgreSQL, RabbitMQ**
  - **Пайплайны разметки данных**: Разработал Airflow-пайплайны, автоматизировавшие разметку данных для задач классификации изображений в R&D-подразделении. **Python, Airflow, Yandex Toloka**
  - **Классификация NSFW-изображений**: Разработал модель обнаружения NSFW-контента, превзошедшую популярные open-source решения на публичных и внутренних бенчмарках, и развернул её в инфраструктуре заказчика. **Python, PyTorch, PyTorch Lightning, TorchServe, Albumentations**
  - **Детекция падения людей**: Разработал модель детекции падения людей на основе адаптированной к домену нейросети поверх open-source модели распознавания скелета. Достиг высокой доли true positive срабатываний при крайне низкой доле false positive на собранных мною бенчмарках и отвечал за полный цикл от сбора и разметки данных до интеграции в продукт. **Python, PyTorch, TensorRT, OpenVINO**
  - **Ускорение модели распознавания лиц**: Ускорил GPU-модель распознавания лиц на 10% без потери точности, применив методы neural architecture search для выбора более эффективного backbone. **Python, PyTorch**

- **Лаборатория методов анализа больших данных** — Москва, Россия
  Research Intern — *янв 2020 – ноя 2020*
- **Deeplight Ventures** — Москва, Россия
  Software Engineer — *фев 2019 – фев 2020*
- **Сбербанк** — Москва, Россия
  Data Science Intern — *авг 2017 – сен 2017*

Персональные проекты

- **k8s.leonid.sh**: Персональная Kubernetes-платформа в Yandex Cloud. Облачные ресурсы — кластер, сеть, managed PostgreSQL, registry, DNS, хранилище секретов — описаны как код в Terraform и разворачиваются Ansible; всё, что работает поверх, доставляется Argo CD из Git. Платформенный слой включает Envoy Gateway с cert-manager и ExternalDNS для ingress и сертификатов, External Secrets для доступов и self-hosted Prometheus, Grafana и Loki для мониторинга, алертинга и логов. На ней живут остальные проекты отсюда. **Kubernetes, Terraform, Ansible, Argo CD, Envoy, Prometheus**
- **Sein zum Tode**: Telegram-бот, предназначенный для того, чтобы помочь пользователю пережить перспективу смерти как явления, которое произойдёт и, более того, **уже происходит** непосредственно с ним. Пользователю предлагается ответить на ряд вопросов, после чего средствами LLM будет приблизительно оценена оставшаяся продолжительность его жизни. Основную ценность представляет ежедневно уменьшающееся число оставшихся дней, которое бот будет регулярно присылать пользователю. Каждый диалог — workflow в Temporal, PostgreSQL — основная база, а Redis хранит чувствительные ответы лишь временно, из соображений приватности. **Python, aiogram, Temporal, OpenAI API, PostgreSQL, Redis**

Другие достижения

- **Rosneft Seismic Challenge**: Международное соревнование по компьютерному зрению в сегментации сейсмических изображений. Занял 2-е место среди 98 команд-участниц (boosters.pro, окт 2019 – дек 2019).
- **Курс System Design**: С отличием завершил продвинутый курс по проектированию распределённых высоконагруженных систем (karpov.courses, янв 2023 – фев 2023).
- **Силовые тренировки и трансформация тела**: Добился долгосрочной рекомпозиции тела: сбросил более 28 кг веса, одновременно набрав значительную мышечную массу, продемонстрировав дисциплину, последовательность и нацеленность тренировок на результат (апр 2024 – н. в.).

Технические навыки

- 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