Vinogradov Sergey

 Barcelona, Spain ✉ vinogradov.so@phystech.edu 📞 +34 657 040 692 **in** biscuitslayer ⌨ BiscuitsLayer

Experience

Beagle, Senior Backend Developer
Barcelona, Spain
Jun 2025 – Ongoing

- Built template system for documents processing with flexible configuration
- Implemented dataset indexing and export functionality with advanced filtering capabilities and multi-format document support
- Developed complex search and highlight logic with persistent annotation storage and retrieval
- Developed ML backend with agentic tools and RAG for assistant application used by lawyers and their clients

SberDevices, Computer Vision Engineer (Master's Diploma)
Moscow, Russia
Apr 2025 – Jun 2025

- Developed an algorithm for detecting damaged QR-codes based on alignment patterns grid
- Improved the binarization stage of the QR-code detection pipeline, which resulted in increased detection stability
- Used graph algorithms to detect corrupted finder and alignment patterns

ReML, Technical Lead
Moscow, Russia
Feb 2025 – Jun 2025

- Managed backend development team with 3 engineers
- Developed a micro-service architecture for processing thermal data from a large number of cameras and sensors, using existing Mobotix backend in C++ and data processing pipelines in Python
- Worked with Grafana API, made functionality for enabling and disabling metrics, creating and deleting dashboards and panels using API calls
- Made a selenium-based tool for interaction with the Mobotix web interface to receive and send different zone parameters to the camera

ReML, Backend Developer
Moscow, Russia
Feb 2025 – Jun 2025

- Developed backend micro-services architecture for processing a large number of documents in the banking sector
- Built a sophisticated document mining system using LLM (OpenAI GPT or Qwen models) for structured entity extraction and compliance checking
- Designed and implemented a complex relational database schema to manage document lifecycles and verification rules
- Built a distributed task processing system, including a worker pool, message queues and automatic failover to ensure high reliability

ReML, Computer Vision Engineer / Backend Developer
Moscow, Russia
Sep 2024 – May 2025

- Developed backend micro-services for PPE detection with low latency
- Added support for multiple RTSP-stream processing using Gstreamer
- Responsible for Spark-based post-processing module to create events for violation detections on camera

ReML, Computer Vision Engineer
Moscow, Russia
Sep 2024 – Feb 2025

- Developed face-recognition module with FAISS database
- Worked with various vision models and MLLM-s, fine-tuned NNs for segmentation and tracking

- Developed an app to get highlights from a hockey game video for a specific player
- Implemented various filters and algorithms to improve recognition accuracy
- Deployed models to production, did inference testing and optimizations

ScaleGen AI, Backend Developer

Moscow, Russia
Oct 2023 – Sep 2024

- Developed LLM inference and fine-tuning tool, with cloud instances setup and scaling according to RPM
- Developed API Gateway setup pipelines using AWS, GCP and Azure
- Developed billing service and price estimation for various cloud computing instances
- Added support for on-prem instances, including all networking setup and security

Huawei, Graphics Engineer

Moscow, Russia
Jul 2021 – Oct 2023

- Built ground-truth raytracer to verify approximations for ray-tracing pipeline
- Improved Vulkan API validation layers
- Added shaders for different materials processing in physical-based rendering algorithms
- Developed Python CI/CD framework for mobile devices testing and performance measurements on multiple devices

Education

MS	**School of Data Analysis, Department of Innovation and High Technology, MIPT + Yandex**, Data Science	Sept. 2023 – Aug. 2025

- **Courses:** CV (2D and 3D), Self-Driving cars, CUDA, MLOps and DevOps courses
- Developed a startup for travellers for efficient route building using AI
- Diploma on the topic "Method for robust decoding of damaged QR codes with nonlinear geometric distortions for use on mobile devices"

BS	**Department of Radio Engineering and Cybernetics, MIPT**, Mathematics and Computer Science	Sept. 2019 – Aug. 2023

- Mostly involved in compiler and architecture simulator development
- Participated in mobile development and AI hackathons
- Diploma on the topic "Mesh Simplification using algorithms and GNNs"

Skills

Science: Math, Linear Algebra, Probability and statistics, Physics and Theoretical mechanics, Algorithms, Graph Algorithms, Concurrency, Deep Learning, Computer Vision, LLM, VLM, Self-driving, Reinforcement Learning

Languages: C++, C, Assembly, Python, Dart/Flutter, PHP, JavaScript

Databases and Storages: PostgreSQL, MongoDB, ClickHouse, Redis, Minio, Amazon S3, FAISS

Big Data and Streaming: Spark, Apache Kafka, RabbitMQ

Networking and Monitoring: NGINX, Caddy, Prometheus, Grafana, Loki, Promtail

Frameworks and Tools: Django, FastAPI, Celery, Bootstrap, Selenium

ML Tools: PyTorch, LibTorch, PyTorch Lightning, Tensorflow, OpenCV, OpenAI API, LangChain, Triton, Pandas, Polars

Graphics: OpenGL, Vulkan, OpenCL, CUDA, RenderDoc, SmartPerf

Robotics: ROS2, Open3D, Carla, Foxglove, Calibration, MPCC

Cloud and Deploy: AWS, GCP, Azure and Azure Functions, Datacrunch, Docker, Terraform, Kubernetes, Ansible, CI/CD

Compilers: Clang, LLVM IR, GNU Bison, GNU Flex