~~Moscow~~ (relocating)
+7 (915) 426 0833
me@sereja.me
sergey-slotin
sslotin

Sergey Slotin

Summary

I am a **computer scientist** specializing in algorithm design, high performance computing, machine learning, and natural language processing, and a **software engineer** with broad experience in systems programming, full-stack web development, data science, and technical management.

I am ideally looking for a long-term industrial research position in a big tech company or a deep tech startup, in either an individual contributor or semi-managerial role with a high degree of autonomy.

Employment

MLOps / Research Engineer (Speech)
Replika May '20 — January '21

— Implemented a speech synthesis service and an on-device voice activity detection model.
— Developed an internal framework for training and serving machine learning models on Kubernetes.
— Moved the ML infrastructure to the cloud, cutting costs and improving availability with autoscaling.
— Maintained data processing and analytics pipelines and improved AI production services monitoring.

Machine Learning Engineer (Image Search)
Yandex April '19 — January '20

— Applied deep transfer learning methods to improve image search ranking quality.
— Worked on improving the trade-off between search quality and monetization.
— Did extensive feature engineering to boost search quality and reduce bias for specific query types.

Deep Learning Engineer (NLP)
Tinkoff July '18 — April '19

— Developed a retrieval-based conversational chatbot, briefly leading a small team.
— Designed a novel transfer learning approach for intent classification.
— Worked on language modeling and ensemble methods for speech recognition.

After Replika, I took a sabbatical to work on personal projects, fill gaps in my knowledge, and write a book (see below).

In the 4-month timespan between Yandex and Replika, I was running a tiny startup trying to build an AI-enhanced messenger for customer support automation. Amidst the COVID panic, unable to apply to accelerators and raise money, I disbanded it and accepted a stable job.

Hard Skills

Domain knowledge	algorithm design, performance engineering, parallel programming, distributed systems, machine learning, natural language processing, operations research, game theory
Programming languages	C, C++, Go, Python, JavaScript/TypeScript, Bash, Scala, Rust, Julia
Infrastructure	Linux, Docker/OCI, Ansible, Slurm, Kubernetes, Istio, Knative, Helm, Prometheus
Low-level	x86, Arm, GCC/LLVM, CUDA, OpenCL, perf, valgrind
Back-end	Flask, FastAPI, Echo, Express, Redis, MongoDB, PostgreSQL, etcd, NSQ, RabbitMQ, Kafka, OpenAPI, gRPC, GraphQL
Build automation	git, svn, Make, Bazel, GitHub Actions, Travis, TeamCity, Jenkins
Machine learning	PyTorch, TensorFlow, NumPy, SciPy, CatBoost, PyTorch Lightning, Transformers
Data engineering	pandas, Dask, Spark, HDFS, Argo, Airflow, jq, Jsonnet, regex, bs4, Selenium
MLOps	KubeFlow, MLflow, DVC, Sacred, Hydra, Triton
Platforms	GCP, AWS, Datadog, Netlify, MTurk, Toloka
Front-end	HTML, CSS/Sass, Vue, SVG, reveal.js, Hugo, Pandoc
Other software	Jupyter, Matplotlib, Bokeh, SageMath, LaTeX, Google Docs suite

Natural languages	English (C2)	Russian (native)	Chinese (HSK4 ≈ B1)
	German (B2)	Spanish (B1)	French (A2)

Research & Writing

I authored or edited most of Algorithmica, a popular Russian-language online computer science textbook spanning over 150 classic algorithms and data structures — as well as several related open-source tools and libraries.

I am currently writing Algorithms for Modern Hardware, a book on high performance computing, and in the process, I have designed and implemented several new state-of-the-art algorithms and data structures, including:

— S-tree, a static data structure for searching in sorted arrays (up to 15x faster than std::lower_bound).
— B− tree, an in-memory search tree (~15x faster than std::set and 4-5x faster than absl::btree).
— SIMD-based algorithms for searching for a value and finding the index of the minimum element in an array at the speed of memory (~3x faster than the previous best approach).
— A ~20% improvement over an existing prefix sum algorithm.
— Yet unpublished algorithms for the all-pairs dense graph shortest paths problem and parsing/printing integers.

Since I am not an academic, I do not need to submit my work to journals, but I do have one academic paper titled "Semi-Supervised Neural Machine Translation with Language Models," which I co-authored as a freshman at MIPT.

Teaching

I have been teaching computer science courses for ~4.5 years, giving over 500 hours' worth of lectures and seminars on a wide range of topics: from algorithm analysis to parallel programming to deep reinforcement learning.

I co-founded and did most of my teaching at Tinkoff Generation, an educational nonprofit that offers free online and in-person courses to undergrads and gifted high school students and currently trains about half the finalists of the Russian Olympiad in Informatics.

I have also helped organize and/or taught at Fintech School (NLP), 1C Educational Center (CS), GoTo School (DL), Universum (DL), l2sch summer school (CS), and Moscow Region ROI training camp (CS), and gave presentations in internal reading groups at MIPT, Tinkoff, Yandex, and Replika.

Education

Moscow Aviation Institute
Computer Science 2019 — 2021

I had an arrangement similar to the one that American college athletes get but for programming competitions instead of sports: I represented the MAI team at the ICPC World Finals (2020), and the computer science department didn't expel me for academic backlog and extensive absences, allowing me to work full-time.

I technically got to the final year with ~3 years' worth of credits, the rest being mainly irrelevant non-CS classes. I saw no point in finishing them just to graduate, especially during the pandemic, so I never did.

Moscow Institute of Physics and Technology
Computer Science 2017 — 2019

As a freshman, I frequently "hanged around" at iPavlov, a MIPT-based deep learning research lab, partaking in their elective courses, deep learning reading group, and some NLP research projects.

I grew to like machine learning much more than college math, and after getting a full-time job on a promising project, I decided to transfer to a less time-demanding place and focus on my career.

Moscow State School 179
Engineering 2015 — 2017

I did competitive programming in high school and achieved a CodeForces rating of 2315, which at the time was roughly equivalent to IOI high silver.

I never managed to qualify for the IOI, but my highest "official" result is the 2-4th place at the Moscow Open Olympiad in Informatics — ahead of 6 of that year's silver medalists.

Before the pandemic, I have also attended:

— DeepHack.Babel (2018), a workshop and hackathon on machine translation (where my team took 2nd place);
— DeepBayes (2018), a summer school on Bayesian deep learning;
— Most CS and DL research seminars at YSDA in 2019 while I was at Yandex;
— Phystech.Start (fall 2019), a MIPT-based entrepreneurship program and a startup accelerator.