Sergey Slotin

~~Moscow~~ (relocating)
me@sereja.me
sergey-slotin
sslotin

Summary

Computer scientist specializing in high performance computing, machine learning, and natural language processing and **software engineer** with broad experience in systems programming, full-stack web development, and data science.

Currently underemployed (part-time teaching and occasionally consulting) and passively looking for a long-term applied research position in a big tech company or a startup with a strong team, in a senior individual contributor or semi-managerial role with a high degree of autonomy.

Industry Experience

Replika (a YC-backed startup)
MLOps / Research Engineer (Speech) 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.

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

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

Tinkoff (a fintech company)
Deep Learning Engineer (NLP) 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 building 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 more steady job.

Technical Skills

Domain knowledge	algorithm design, performance engineering, parallel programming, distributed systems, machine learning, natural language processing, game theory, operations research
Programming languages	C, C++, Go, Python, JavaScript/TypeScript, Bash, Scala, Rust, Julia
Infrastructure	Linux, Docker, Kubernetes, Knative, Helm, Nginx, Istio, Prometheus, Ansible, Slurm
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
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
SVC & build tools	git, svn, Make, Bazel, GitHub Actions, Travis, TeamCity, Jenkins
Platforms	GCP, AWS, Datadog, Netlify, MTurk, Toloka
Front-end	HTML, SVG, CSS/Sass, Vue, D3.js, reveal.js, Hugo, Pandoc
Other software	Jupyter, Matplotlib, Bokeh, SageMath, LaTeX, Google Docs suite, Inkscape, GIMP

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 working on Algorithms for Modern Hardware, a book on high performance computing.
In the process, I have pushed the state of the art in several fundamental algorithmic problems, including:

— S-tree, a data structure for faster binary search (~15x faster than `std::lower_bound`);
— a B-tree implementation (~15x faster than `std::set` and ~5x faster than `absl::btree`);
— SIMD-based algorithms for searching for a value (~10x), argmin (~3x), and prefix sum (~1.2x) on an array;
— a word-sized integer factorization algorithm (~3x);
— yet unpublished algorithms for parsing/printing integers and the all-pairs dense graph shortest paths problem.

Since I am not in academia, 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" that I co-authored as a freshman at MIPT.

I also have ~5 papers' worth of yet-to-be-properly-published research on rating systems and social deduction games.

Teaching

I have been teaching computer science courses for 5 years, giving over 500 lectures and seminars on topics from algorithm analysis to parallel programming to deep reinforcement learning.

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

I have also helped organize and/or taught at many other Moscow-based educational institutions, summer schools, training camps, and internal reading groups at tech companies and universities and co-organized and contributed problems for minor programming competitions and hackathons. More recently, I switched to mainly speaking at tech conferences, giving presentations at CppCon, Meeting C++, Performance Summit, and C++ Zero Cost Conf.

Education

Moscow Aviation Institute
Computer Science 2019 — 2021

I represented the MAI team at the ICPC World Finals (2020) and—in an arrangement similar to the one that American college athletes get but for scientific olympiads instead of sports—the computer science department tolerated my extensive absences and academic backlog, allowing me to work full-time.

I got to the final year with ~3 years' worth of credits, the rest mostly being 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 (Russia's hardest university)
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 (ranked #1 technical school in Moscow in my graduation year)
Engineering 2015 — 2017

I did competitive programming in high school and achieved a CodeForces rating of 2315, at the time 3rd-highest in Russia and 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.