

ANDRII SEMENOV

1005, Lausanne, Switzerland

✉ [Email](#) [in](#) [Profile](#) [G](#) [GitHub](#) [Globe](#) [Website](#) [G](#) [Scholar](#)

EDUCATION

École polytechnique fédérale de Lausanne

MSc in Data Science

Sep 2024 -

- [School of Computer and Communication Sciences](#)
- [Machine Learning and Optimization Laboratory](#)

Moscow Institute of Physics and Technology

BSc in Applied Mathematics and Physics

Sep 2020 - Aug 2024

- [Landau Phystech School of Physics and Research](#)
- [Chair of Problems of Physics and Astrophysics](#)
- Affiliated with [Yandex chair of Data Analysis](#)
- Advisor: [Aleksandr Beznosikov](#)
- Thesis: "[Contrastive Learning for Enhancement of Model Interpretability in Computer Vision](#)"

WORK EXPERIENCE

Research Assistant | [Machine Learning and Optimization Laboratory, EPFL](#)

Aug 2024 -

- LLM pretraining, Optimization, Training Dynamics
- Head: Professor [Martin Jaggi](#)
- Resulting Papers: [\[11\]](#), [\[10\]](#), [\[9\]](#), [\[8\]](#), [\[7\]](#), [\[6\]](#)

Visiting Researcher | [Machine Learning Department, MBZUAI](#)

Feb 2025 - Apr 2025

- Head: Professor [Martin Takáč](#)

Research Student | [MIPT-Yandex Fundamental Research Laboratory](#)

Jul 2023 - Oct 2024

- Machine Learning and Optimization
- Head: PhD [Aleksandr Beznosikov](#)
- Resulting Papers: [\[5\]](#), [\[2\]](#)

Research Student | [Laboratory of Mathematical Methods of Optimization, MIPT](#)

Jul 2023 - Oct 2024

- Optimization
- Head : Professor [Alexander Gasnikov](#)
- Resulting Papers: [\[3\]](#), [\[1\]](#)

Deep Learning Engineer | Huawei-MIPT research group

Nov 2023 - Oct 2024

- Deep Learning and Reinforcement Learning
- Head : Professor [Roland Hildebrand](#)
- Resulting Papers: [\[4\]](#)

Research Student | [Lab of Fundamental and Applied Research of Relativistic Objects](#)

Nov 2022 - Apr 2024

- Theoretical Physics, Astrophysics
- Head : DSc [Elena Nokhrina](#)

Research Physicist | [P.N. Lebedev Physical Institute](#)

Nov 2022 - Jul 2023

- Theoretical Physics, Astrophysics

SKILLS

Stack Python, PyTorch, TensorFlow, JAX, C++, C#, \LaTeX , SQL, Git, Linux, macOS
Language English – C1, Russian – native, Ukrainian – native
Hobbies Swimming, Football

RESEARCH INTERESTS

Large-Scale Optimization, Natural Language Processing, Federated Learning and applications of Stochastic Optimization in Deep Learning

PUBLICATIONS

Enhancing LLM Training via Spectral Clipping

Mar 2026

Preprint

Xiaowen Jiang, **Andrei Semenov**, Sebastian U. Stich

- [arXiv](#)
- [Code](#)

Adaptive Regularized Newton Methods with Inexact Hessian

Dec 2025

International Conference on Computational Optimization, 2025 (Oral Presentation)

Alexander Shestakov, Nail Bashirov, **Andrei Semenov**, Alexander Gasnikov, Martin Takáč, Aleksandr Beznosikov, Dmitry Kamzolov

- [arXiv](#)
- [PDF](#)

Apertus: Democratizing Open and Compliant LLMs for Global Language Environments

Sep 2025

Preprint

Swiss AI Initiative, Apertus Team (contributor)

- [arXiv](#)
- [Technical Report](#)
- [Code](#)
- [Open-sourced models](#)

Benchmarking Optimizers for Large Language Model Pretraining

Sep 2025

Preprint

Andrei Semenov, Matteo Pagliardini, Martin Jaggi

- [arXiv](#)
- [Code](#)
- [Wandb Logs](#)
- [EurIPS 2025 Poster](#)

Gradient-Normalized Smoothness for Optimization with Approximate Hessians

Jun 2025

ICLR 2026

Andrei Semenov, Martin Jaggi, Nikita Doikov

- [arXiv](#)
- [Code](#)

Sign Operator for Coping with Heavy-Tailed Noise: High Probability Convergence Bounds with Extensions to Distributed Optimization and Comparison Oracle

Feb 2025

Preprint

Nikita Kornilov, Philip Zmushko, **Andrei Semenov**, Alexander Gasnikov, Aleksandr Beznosikov

- [arXiv](#)
- [PDF](#)

Just a Simple Transformation is Enough for Data Protection in Vertical Federated Learning

Dec 2024

Preprint

Andrei Semenov, Philip Zmushko, Alexander Pichugin, Aleksandr Beznosikov

- [arXiv](#)
- [Code](#)

Mixed Newton Method for Optimization in Complex Spaces

Jul 2024

Preprint

Nikita Yudin, Roland Hildebrand, Sergey Bakhurin, Alexander Degtyarev, Anna Lisachenko, Ilya Kuruzov, **Andrei Semenov**, Mohammad Alkousa

- [arXiv](#)
- [PDF](#)

Clipping Improves Adam-Norm and AdaGrad-Norm when the Noise Is Heavy-Tailed

Jun 2024

ICML 2025

Savelii Chezhegov, Yaroslav Klyukin, **Andrei Semenov**, Aleksandr Beznosikov, Alexander Gasnikov, Samuel Horváth, Martin Takáč, Eduard Gorbunov

- [arXiv](#)
- [Code](#)
- [ICML 2025 Poster](#)

Sparse Concept Bottleneck Models: Gumbel tricks in Contrastive Learning

Feb 2024

Preprint

Andrei Semenov, Vladimir Ivanov, Aleksandr Beznosikov, Alexander Gasnikov

- [arXiv](#)
- [Code](#)

Bregman Proximal Method for Efficient Communications under Similarity

Nov 2023

International Conference on Computational Optimization, 2024 (Oral Presentation)

Aleksandr Beznosikov, Darina Dvinskikh, Dmitry Bylinkin, **Andrei Semenov**, Alexander Gasnikov

- [arXiv](#)
- [PDF](#)
- [ICOMP 2024 Poster](#)

PROJECTS

Improved Megatron-LM | a codebase for large-scale training and inference

Feb 2025 –

Swiss AI Initiative & EPFL, Machine Learning and Optimization Laboratory

[Project Link](#)

- This repository is designed to help researchers create reproducible experiments at scale
- Allows training of up to 70B models
- Inspired by the [NVIDIA](#) codebase and its [variant from Swiss AI](#)

Learning@Scale & llm-optimizer-benchmark | nanoGPT-like codebases for pretraining

Oct 2024 –

EPFL, Machine Learning and Optimization Laboratory

[Project Link](#)

- These repositories are designed to help researchers create reproducible experiments
- Designed for the small-scale experiments ($\leq 8B$ models) with the DDP paradigm
- Inspired by the [epfml/llm-baselines](#) codebase and the [nanoGPT](#) benchmark
- [llm-optimizer-benchmark](#) is a supplementary codebase for the [paper](#)

Llama-LoRA project | Natural Language Processing, Transformers

Jul 2023

MIPT

[Project Link](#)

- Fine-tuned 13B and 7B models on a custom dataset containing my Telegram chats
- Pushed my models to HuggingFace hub. Where they got 10000+ downloads! [HuggingFace Link](#)

TALKS

- 9 October 2024, [MLO Group Meeting](#). Talk on the "Defense against Feature Reconstruction attacks" [[slides](#)]
- 26 March 2024, MIPT-Yandex Optimization Seminar. Talk on "Model Reconstruction Attacks" [[video](#)]
- 12 March 2024, MIPT-Yandex Optimization Seminar. Talk on "Concept Bottleneck Models" [[video](#)]

TEACHING

Moscow Institute of Physics and Technology

Teaching Assistant

Jan 2024 - May 2025

- Spring 2025: Stochastic Analysis (remote mentor)
- Autumn 2024: Machine Learning. Part of the [MSAI](#) team, course [repository](#)
- Spring 2024: Reinforcement Learning. Owner of the course [repository](#)

REVIEWING

- **NeurIPS**: 5 papers in 2025
- **ICLR**: 3 papers in 2024, 5 papers in 2025
- **ICOMP**: 4 papers in 2025

Last updated on Mar 17, 2026