Stepan Shabalin

https://neverix.io/

Email : sshabalin3@gatech.edu

Mobile : +1 786 836 7862

EDUCATION

- **Georgia Institute of Technology** — Atlanta, GA
  Bachelor of Science in Computer Science; GPA: N/A — *Aug. 2023 – Current*

- **School 4 of Cousteau** — St. Petersburg, Russia
  High School; GPA: 3.82 (4.77/5) — *Sep. 2021 – May 2023*

EXPERIENCE

- **MIT Reality Hack** — Online
  Web Tools Team — *Apr 2021 - Current*
  - **Hardware checkout system**: Built workflow for making requests for AR and VR hardware during the event thorugh scanning QR codes and keeping track of ownership using Next.js and Firebase.
  - **Backend maintenance**: Helped keep the backend up under load during the event by adding caching.
  - **Exports and judging system**: Exported hackathon projects from previous years from Devpost and helped set up projects for judging.

- **Sberbank** — St. Petersburg
  Intern — *Aug 2021*
  - **Recommendation systems research**: Researched transformers for recommendation systems with reinforcement learning.

- **Deepfloyd** — Online
  Intern — *Aug 2022 - Dec 2022*
  - **Autoencoders for latent diffusion models**: Ran experiments comparing different ways of training variational autoencoders for diffusion models using PyTorch and Taming Transformers.
  - **Latent 3D Diffusion models**: Researched generation of 3D models using latent diffusion models.

- **Supervised Program in Alignment Research** — Online
  Mentee — *Sep 2023 - Dec 2023*
  - **Research**: Researched sycophancy in language models under the guidance of Aaron Scher.

- **ML Alignment & Theory Scholars** — Online
  Mentee — *May 2024 - Aug 2024*
  - **Training phase**: Trained sparse autoencoders and used them for feature steering under the guidance of Neel Nanda and Arthur Conmy.
  - **Research program**: Researching the use of sparse autoencoders for interpretability.

HONORS/AWARDS

- **Atlas Fellowship**: Finalist; $1,000 prize. Program aiming to "transmit a culture inspired by scientific inquiry, with the ambition to tackle neglected global challenges" to high school students. Over 13,000 applicants, $< 3\%$ acceptance rate.

- **Non-Trivial Fellowship**: Finalist. Similar program for high school students providing guidance and funding.

CONFERENCES

- **AWE XR 2020**: Presented MIT Reality Hack project draft360.
- **NeurIPS 2023**: Presented spotlight paper MindEye.
- **IEEE SaTML**: Presented 3rd place solution to RLHF Trojan Competition.

Projects

- **AAA-py**: Rewrote the build system for Arcane Algorithm Archive, an online book on algorithms. It has been used since 2018.

- **draft360**: Built the webXR frontend for draft360, a virtual and augmented reality storyboarding tool. Won a $3500 prize at the 2020 MIT Reality Hack hackathon.

- **Golden Transformer**: Fine-tuned dozens of transformers on the Russian SuperGLUE benchmarks and combined them into an ensemble to achieve state-of-the-art performance in 2021. This project won a 500,000 ₽ prize in the National Technology Olympiad in AI.

- **Video generator**: Collaborated on a project that generates videos from text descriptions using generative AI. Won a 500,000 ₽ prize in the AI International Junior Contest.

- **3D generation**: Published early (January-December 2022) Colab notebook implementations of Dream Fields, DreamFusion and VectorFusion on Twitter.

- **MindEye**: Applied retrieval from a web image dataset for improving image generation quality for MindEye, a system for generating images from fMRI brain scans. Spotlight paper at NeurIPS.

- **micrlhf/saex**: Developed a codebase for quantized model parallel inference of GGUF models in Jax. Trained sparse autoencoders for Phi-3 Mini and Gemma 2B.