

# Artem Petrov

Berlin, Germany / Remote / Relocation • +995 (591) 062-551 • artem99petrov@gmail.com • GitHub: ArtemPt239

## SKILLS

- LLMs:** Prompt Engineering, Agent harness design, Tracing, LangChain, RAG, Evaluations  
**Data science:** Experimental design, Machine Learning, Deep Learning with Pytorch, Computer Vision (segmentation with Unet), Data augmentation, Data visualization  
**Software development:** Python (6 years) (data science stack; PyTorch; FastAPI; SQLAlchemy, Plotly:dash), SQL, Git, Docker, Bash, Java/Kotlin for Android (1 year), C++ (6 months)  
**Other:** advanced Linux user, advanced Cursor user, L<sup>A</sup>T<sub>E</sub>X, typst

## EXPERIENCE

- Project Lead & AI Safety Researcher @ Palisade Research** Jul 2024 – Present
  - Built demos to raise awareness of AI Safety issues, reaching millions of people
  - Evaluated evals gap in AI cyber capabilities, achieving state-of-the-art results, and explored ways to fix it
  - "These results are great!" - RAND team about my work during Palisade-RAND collab
- Python Developer & Prompt Engineer @ We Are Volt** Feb 2024 – Aug 2024
  - Built, tested, and deployed a multi-staged= RAG pipeline for Q&A and Analytics over a custom CMS containing visual, textual and numerical information sources
- Data Scientist & Python Developer @ Tentakel** Jan 2023 – Feb 2024
  - Developed pipelines for automatic data analysis and document Q&A with LLMs and wrapped them in a REST API
  - Deployed LLM-based products in the form of Telegram bots and Slack apps
- Data Science Researcher @ IRA-Labs [part-time]** Jul 2021 – Jun 2022
  - Trained UNets for organ segmentation on CT images
- Data Science Intern @ Russian Academy of Sciences, IITP** Aug 2020
  - Investigated viability of implementing attention in Simple Graph Convolution neural net architecture
- Python developer** Jun 2020 – Jul 2020
  - Developed a web dashboard to represent internal data with Plotly|dash
- System administrator [part-time]** 2017 – 2024
  - Set up, maintained, and migrated Linux-based mail servers

## PUBLICATIONS

- GPT-5 at CTFs: Case Studies From Top-Tier Cybersecurity Events** Nov 2025
  - We show GPT-5 outperforming 90% of human hackers on one of the hardest public CTF competitions
- Evaluating AI cyber capabilities with crowdsourced elicitation** May 2025
  - We host AI tracks on CTF competitions and show AI outperforming 79% of human hackers
- Biollama: testing biology pre-training risks** Feb 2025
  - We collaborate with RAND to find if adversaries can fine-tune LLMs for use as bio lab assistants
- Hacking CTFs with Plain Agents** Dec 2024
  - We achieve SOTA on a CTF benchmark with a simple agent harness

## LECTURES & TALKS

- Speaker @ AI Governance workshop at AAAI** Jan 2026
  - Spoke about current state of AI Security and its differences from traditional cybersecurity
- Keynote Speaker @ Zurich AI Safety Day** Sep 2025
  - Presented "Getting started in AI safety" talk to an audience of 150 people
- Lecturer @ Lalambda School** Aug 2023  
A biyearly summer school on advanced CS topics
  - Taught a 16-hour course on the latest advancements in NLP with a deep dive into the Transformer Architecture
- Seminarist @ Tbilisi Paper Reading Club** 2023
  - Led a couple of dozen seminars on fundamental deep learning papers

## EDUCATION

- ARENA (Alignment Research ENgineer Accelerator)**, London, UK Apr 2025 – May 2025
  - An upskilling program for AI safety researchers designed by engineers from Google Deepmind
- Yandex School of Data Analysis**, Moscow, Russia 2022 – 2023

- Part-time masters-level private program in data science and software engineering

**Moscow Institute of Physics and Technology**, Moscow, Russia 2017 – 2022

BS, Applied Mathematics and Physics: Computer Technologies and Intellectual Data Analysis

  - GPA: 3.9/4, **top-2%**
  - Top-1 university in Russia in CS according to THE university ranking

**Physics and Mathematics Lyceum №239**, Saint-Petersburg, Russia 2013 – 2017

  - Every year from 2015 to 2020, the Lyceum was recognized as the best school in Russia

**LANGUAGES**      **English:** C1 \ Full Professional Proficiency  
                      **Russian:** Native language

<b>PET PROJECTS</b>	<b>ns-scheduler</b> A tool to schedule the work time of applications deployed to the Kubernetes cluster for development/testing purposes to not waste resources when no one is using them <b>rainbow-track</b> GNOME shell extension for smooth time tracking on Linux with Toggl Track <b>Gold Medal Awardee</b> at National and International Physics Olympiads Judge at the finals of the National Physics Olympiad	2015 – 2017 2018 – 2019
---------------------	---	----------------------------