

# Ivan Pavlov

Lausanne, Switzerland | +41 76 785 91 02 | [pavlov200912@gmail.com](mailto:pavlov200912@gmail.com)

## EDUCATION

---

### EPFL, Lausanne, Switzerland

*Master of Science in Data Science*

Sep. 2024 – Dec. 2026

**GPA 5.83/6.0**

### Saint Petersburg State University, Russia

*Bachelor of Science in Computer Science, Faculty of Math and Computer Science*

Sep. 2018 – June 2022

**GPA 4.91/5.0 (with honors)**

## RESEARCH

---

### Finding Explanations for Emergent Misalignment in LLMs

Feb. 2026 – Present

*Research Project Student, Learning & Adaptive Systems (LAS) Group, ETH Zurich*

- Evaluating mechanistic explanations for Emergent Misalignment in fine-tuned LLMs using linear probes, SAE features, and activation steering;
- Supervised by Cynthia Xin Chen and Lukas Fluri (Prof. Andreas Krause / Prof. Florian Tramèr groups).

### Post-training for SwissAI (Apertus): SFT → DPO

Feb. 2025 – Jun. 2025

*Research Project Student, CLAIRE Lab (EPFL)*

- Reproduced the Tulu3 post-training pipeline on Llama 3.1 8B and adapted it to SwissAI **Apertus** (8B/70B): Supervised Fine-Tuning (SFT) followed by Direct Preference Optimization (DPO).
- Identified that DPO degraded MMLU despite clean training metrics, surfacing a gap between optimization signal and downstream capability.
- Extended evaluation infrastructure `olmes` ([github.com/swiss-ai/olmes](https://github.com/swiss-ai/olmes)); added vLLM inference with tensor parallelism, standard benchmarks (MMLU, GSM8K, HumanEval, IFEval, BBH, TruthfulQA, PopQA), and automated SLURM job generation.
- Implemented and productionized a TRL/DeepSpeed-based DPO trainer in the SwissAI codebase ([github.com/swiss-ai/swiss-alignment](https://github.com/swiss-ai/swiss-alignment)); with full W&B logging and reproducibility scripts.

## PROFESSIONAL EXPERIENCE

---

### Oracle Labs, Research Intern

*LLMs & AI Agents for Schema Mapping*

Aug. 2025 – Feb. 2026

Zurich, Switzerland

- Built an LLM agent (tools-based architecture) for mapping external customer schemas to Oracle's internal schema, improving over the previous baseline; set up evaluation infrastructure for the associated RAG pipeline.

### JetBrains, ML Engineer

*JetBrains AI Team*

Sep. 2022 – Sep. 2024

Paphos, Cyprus

- Contributed to Mellum LLM and code completion system as part of the Evaluation team; owned A/B testing metrics, reporting, and data processing.
- Developed a unified A/B testing pipeline across multiple languages and IDEs, enabling systematic comparison of completion models.
- Trained and deployed CatBoost models for IDE search ranking, improving key ranking metrics by 15%.
- Optimized metric/log collection algorithms for online experiments, reducing storage usage by 30%.
- Implemented documentation generation and refactoring features in the JetBrains AI Assistant plugin.

### Huawei Research, Junior ML Engineer

*ML Research Team*

May 2022 – Aug. 2022

Saint Petersburg, Russia

- Designed a code-centric document retrieval system (DPR, RAG, Transformer encoders), doubling PR@1 over a strong Elasticsearch baseline for Python documentation search.

### JetBrains Research, Intern

*Machine Learning for Software Engineering Lab*

Sep. 2020 – May 2022

Saint Petersburg, Russia

- Built an LSTM-based code-comment inconsistency detection plugin and a 4000+ Java project dataset, refined via a custom Toloka crowdsourcing and active learning pipeline.

## TEACHING & VOLUNTEERING

---

**Teaching Assistant**, Discrete Mathematics I, Neapolis Univ. Pafos (Prof. A. Kulikov)

Oct. 2023 – Sep. 2024

**Instructor** | Olympiad Mathematics Club for Schoolchildren, Paphos, Cyprus

Sep. 2022 – May 2023