

# Andrii Nikolaiev, PhD

📍 Cambridge, UK

☎ (+44) 7768 773026

✉ adnikolaiev@gmail.com

🌐 linkedin.com/in/andynik

🌐 andynik.github.io

## SUMMARY

AI researcher with a PhD and 5+ years in AI/ML, specialising in LLM evaluation, benchmarking, and experimentation design. EdTech founder with 9+ years in building STEM initiatives, educational programmes, and startup operations.

## EDUCATION

### Visiting Research Student

NLIP Group, University of Cambridge (2023–2025)

### Ph.D. in Computer Science

Taras Shevchenko National Univ. of Kyiv (2020–2025)

### B.Sc. & M.Sc. in Computer Science

Taras Shevchenko National Univ. of Kyiv (2014–2020)

## GRANTS & AWARDS

- Emergent Ventures Fellow, Mercatus Center, USA (2023–2026)
- Fully funded research fellowship, University of Cambridge, UK (2023–2025)

## SKILLS

- **Programming:** Python, R, C/C++, SQL,  $\LaTeX$
- **LLM Evaluation & Research:** Benchmarking, CoT, LLM-as-a-Judge, prompt engineering, human evaluation
- **ML Infrastructure:** PyTorch, Hugging Face, vLLM, llama.cpp, CUDA, Docker, GCP, Linux
- **Program Management:** Curriculum design, budgeting, team leadership, stakeholder partnerships
- **Languages:** English (fluent), Ukrainian (native)
- **Interests:** Half-marathon (PB: 01:31:16), football, chess, yoga

## RESEARCH & PROFESSIONAL EXPERIENCE

### Expert Project Manager, Mercor (APEX Research)

Feb 2026 – Present

- Designed rubrics and ran experiments evaluating AI models on economically meaningful professional tasks across domains.
- Analysed AI agent behaviour via MCP tool usage, computer user actions (CUA), browser user actions (BUA).
- Collaborated with stakeholders and domain experts across industries.

### Researcher in NLP, University of Cambridge

Apr 2023 – Present

- Developed novel symbolic benchmarks for combinatorial reasoning, revealing reasoning gaps (up to 30%) across frontier LLMs.
- Deployed and evaluated open-source and proprietary LLMs on HPC clusters and cloud VMs; led human evaluation studies with 40 participants.

### Researcher – LLM Benchmarking, Osiris-AI

Aug – Sep 2025

- Architected a distributed multimodal pipeline for OCR and segmentation of historical handwritten records, achieving a 15% accuracy gain via custom RAG and prompt engineering.

### Ph.D. Researcher in CS, Taras Shevchenko National Univ.

2020 – 2025

- Conducted large-scale benchmarking and analysis of LLM reasoning using over 1M Chain-of-Thought examples.
- Implemented LLM-as-a-Judge frameworks, reducing evaluation costs while maintaining strong alignment with expert judgements.

## SELECTED PUBLICATIONS

- Neural Network Methods for Selecting and Generating Synthetic Variations of Combinatorial Problems, *Cybernetics and Systems Analysis*, Springer Nature, 2025.
- Can Language Models Rival Mathematics Students? Evaluating Mathematical Reasoning through Textual Manipulation and Human Experiments, *ACL RR*, arXiv preprint, 2024.

## ADDITIONAL EXPERIENCE

### Co-founder & CEO, Kvanta.STEM EdTech Startup

2016 – Present

- Built a national STEM programme with 1,000+ learners, 40+ instructors, and \$150K+ in secured funding.
- Designed Olympiad-level curricula in mathematics, programming, and science; led boot camps for up to 200 participants (online & in-person).

### Program Management Experience

2019 - 2022

- Worked as a project manager at Roosh Invest. Group, building partnerships with academic institutions, educational organisations, and stakeholders; coordinated online events for 200+ participants.
- Managed national educational programmes and executed \$70K budgets for nurturing STEM talent at NGO Kontora Pi.

### Teaching & Outreach

- Assisted seminars at universities (ML for Real Data at the University of Cambridge, databases and cloud computing at the Kyiv National University).
- Served as a national jury member on mathematical and AI competitions.