Amir Nuriyev

amir.ravnur@gmail.com | LinkedIn | GitHub | Google Scholar | anuriyev.com

Education

MBZUAI — Abu Dhabi, UAE
B.S. in Artificial Intelligence — *Expected May 2029*

Experience

Research Scholar — June 2025 – October 2026
ML Alignment & Theory Scholars (MATS), mentored by Gabriel Kulp (RAND, ISL) — *Berkeley, CA*

- Developed an inversion attack that recovers 91.2% of input tokens from MoE models' expert selections
- Built an open-source toolkit for side-channel capture from Python, CUDA, and model workloads
- Technical mentor to all research fellows at Intelligence Security Laboratories, Gabriel Kulp's RAND spin-out
- Onboarded fellows from Heron, SPAR, Pivotal, and MATS and gave day-to-day technical direction on their projects

Research Fellow — January 2026 – May 2026
Heron AI Security — *Remote*

- Led a 5-person research team of cybersecurity professionals and PhDs
- Showed that power traces reveal which MoE experts are active, extending routing leakage to hardware side channels

Research Fellow — February 2025 – May 2025
Supervised Program for Alignment Research — *Remote*

- Originated a new research direction: RF and power side-channel attacks on MoE expert routing
- Fine-tuned MoE LLMs to emit side-channel signals and built classifiers to detect them

Data Science Intern — June 2023 – August 2023
Kazakhtelecom — *Almaty, Kazakhstan*

- Calculated bill relief for customers affected by the Abay region wildfires (600K+ people affected)
- Built ML models to select the optimal mobile plan offer for 100K+ customers

Selected Works

A. Nuriyev, G. Kulp. Expert Selections In MoE Models Reveal (Almost) As Much As Text (Paper, 2026)
ICLR Workshop on Trustworthy AI, SAGAI Workshop @ IEEE S&P (Oral)

A. Nuriyev. Reverse-engineering Tinker's Dimuon (Blog, 2026)

Achievements

Kazakhstan IOI Team Selection Training (top 10 nationally)
Codeforces Master (top 3%)
International Zhautykov Olympiad (silver medal, top 70/300)

Talks & Media

Compute Verification Via Side Channels, MTS (Interview, 2026)
Workload Verification Using Side Channels, AViD Workshop @ IEEE S&P (Talk, 2026)

Technical Skills

Languages: C/C++, Python, SQL
Developer Tools: Git, Docker, SLURM
Libraries: PyTorch, Transformers, Pandas, NumPy, Matplotlib