Ansar Zeinulla

ansar.zeinulla@nu.edu.kz | linkedin.com/in/ansar-zeinulla
github.com/ansarzeinulla | huggingface.co/ansarzeinulla | ansarzeinulla.github.io | ORCID 0009-0000-7763-1546

Education

Nazarbayev University — Astana, Kazakhstan
Bachelor of Science in Computer Science — *Aug. 2024 – Jun. 2028 (Expected)*
- **GPA 3.87/4.00, ranked 1st of 221** in the cohort; Dean's List three times; full state scholarship. Taught in English.

Tomorrow School (alem.ai) — Astana, Kazakhstan
Project-based programming school, first graduate — *Jan. 2024 – Jul. 2026*

Research & Engineering

Low-precision inference kernels (llama.cpp, amxpy) — Astana, Kazakhstan
Independent — *2025 – Present*
- `llama.cpp` PR #27589: 8×8 `i8mm` GEMM for ternary `TQ2_0` weights, prompt processing **73.4 → 131.4 tok/s (+79%)** on an M2 Pro.
- Found the patch's 13% regression under Accelerate BLAS, fixed it in the model loader and reworked the tests with the Arm reviewer.
- `amxpy`: ternary GEMM on Apple's undocumented AMX coprocessor, **3.64 TOPS** (2.65× fp32 `sgemm`); slower at batch 1, and documented why.

Strong solution of Bestemshe (with M. Manassov) — Manuscript in preparation
Retrograde analysis; tablebase; neural evaluation — *2025 – Present*
- Solved all 125,698,242,527 positions of a Kazakh mancala: a forced win for the second player.
- Public **8.96 GB** tablebase, any position read with one HTTP range request (Awari, 2002: 778 GiB on 72 nodes).
- A 17.2M-parameter network keeps the value in 99.62% of moves yet loses 54.3% of winnable games; its policy head, 12.7%.

State-space complexity of Togyzkumalak (sole author) — ICGA Journal, revision under review
C++ engine, 10^9 *simulated games* — *2025 – Present*
- Upper bound of 1.51×10^{25} on the state space; game-tree complexity $\approx 10^{105}$ from 10^9 playouts.
- Alpha-beta engine: 3.51M nodes/s native, 2.42M in WebAssembly. Fixed a transposition-table defect found in review; added a regression test.

NogaiLLM, language model for an endangered Turkic language (sole author) — Manuscript in preparation
Corpus; LoRA on MLX — *2025 – Present*
- 2.35M-word corpus; Qwen2.5-1.5B adapted on a laptop, held-out bits per byte 2.04 → 1.09.
- Found train/validation leakage in my own evaluation; withdrew a journal submission and rebuilt the split.

Research Assistant, SOILEU spoken corpus — Nazarbayev University
Under Prof. Andrey Filchenko — *Jun. – Aug. 2025*
- One of 10 student researchers processing audio/text data of Central Asian languages into a multimedia corpus.

Publications & Talks

A. Zeinulla. *Combinatorial State-Space and Empirical Game Tree Complexity of Togyzkumalak: A Billion-Game Analysis.* CMAI 2026 Book of Abstracts, pp. 51–53, ISBN 978-601-08-6470-2. Talk, 9 Sep 2026.
Full paper: second round of review, *ICGA Journal*.

Honors

XVI Presidential Olympiad in natural-mathematical sciences, 2nd-degree diploma, Ministry of Education (2023)
International Mathematics Competition (IMC), Bronze Medal, individual contest, Durban (2019)

Technical Skills

Languages: C++17, Python, WebAssembly (Emscripten), Bash, SQL, LaTeX
Systems: Apple AMX, ARM NEON / `i8mm` kernels, multi-threading, cache-aware data layout, ZSTD block compression, retrograde analysis
Machine Learning: PyTorch, MLX / `mlx-lm`, LoRA fine-tuning, ternary and low-bit quantization, alpha-beta search, evaluation design
Languages spoken: Kazakh (native), Russian (fluent), English (fluent, English-medium degree)