FOR IMMEDIATE RELEASE

Rule-Based Information Engine Holds 54-Megabyte Data Array as a Fixed 32-Byte Rule

Storing the rule instead of the output changes what machine-generated data costs to keep, and the saving grows with the size of the array.

DUBAI, UAE — August 10, 2026 — The Knowledge Triangle Route (KTR) project, an independent research framework based in the United Arab Emirates, today published a reproducible benchmark showing how a 54-megabyte data array can be stored and reconstructed using a single 32-byte mathematical rule. A US provisional patent application for the method was filed in June 2026. The Information Engine (INEN) keeps the mathematical rule that generates an array rather than a snapshot of it, reconstructing the original data bit-for-bit on demand.

Developed by independent inventor Roman Burtsev, INEN is evaluated against production-grade differencing followed by LZMA, the Lempel-Ziv-Markov chain algorithm that underlies the standard baseline used in systems such as Apache Parquet (DELTA_BINARY_PACKED), FLAC, and Gorilla.

“We evaluate against delta encoding, not against general-purpose compressors fed ASCII text. The delta baseline is what cloud engineers actually run in production, so that is the baseline worth publishing,” said Roman Burtsev, inventor of the Information Engine.

The Rule Stays Fixed While the Baseline Grows

The stored rule is 32 bytes at every array length, while the delta coder must record state changes for every element, so its output grows from 108 bytes at a thousand elements to 7,696 bytes at 6.5 million, widening the gap from 3-fold to 240-fold across the range. At that size, the raw array occupies 54,796,312 bytes and the rule occupies 32 — the same 32 bytes that hold a thousand.

The advantage becomes structural on machine-generated data such as simulation output, synthetic AI datasets and high-frequency logs, which is globally deterministic but locally random. On the output of a linear congruential generator, a 56-byte rule opens a 7,881-fold gap over the same delta baseline; on Fibonacci mod p, a 36-byte rule opens a 15,440-fold gap. Differencing cannot close these gaps by construction, and a stronger compressor would not help either: LZ-family methods already sit at the entropy rate of a stationary source, while generated data carries high entropy at low generative complexity.

Three Things a Compressor Cannot Do

The method also recovers data that is missing: when 500 values were deleted from a series, it restored 499 of them exactly and got none of them wrong. A compressor cannot do this, because it returns only what it stored.

Where there is no rule in the data, INEN returns a NO_RETURN status instead of a result. No compressor can give that answer, because a compressor always returns something. A system that reports the absence of structure does not pass an invented number downstream.

The compression ceiling is computable before compressing, from a single measurement of the stream. For integer arithmetic it is unbounded, while for chaotic dynamics it is finite and can be named before the work begins.

“A ratio can be disputed. A ceiling named before the work starts is a commitment,” Burtsev said.

The ceiling is 64 bits per sample divided by the Lyapunov tariff of the stream: a single measurement, taken before any compression runs.

Because the rule is found once while the saving accrues on every read, access frequency decides the economics. According to the project's own estimate, a cold archive read once a year never pays back, while

a hot cache at a million reads a year pays back within days, placing the method in active data paths rather than long-term storage.

Built to Be Checked

Every quantitative result is produced by a terminating script that reproduces the numbers bit-for-bit, and every comparison carries both baselines, the naive one and the standard one. The standard one is what is published. The full benchmark, codebase, and execution logs are deposited on the Open Science Framework (osf.io/kjpvg), allowing any external body to confirm the hashes against the manifests and rerun the work independently.

Machine-generated data now grows faster than anything written by people, and it is printed by rules. Stored as files, it preserves the trace instead of the cause. For data centre and cloud operators, the practical question shifts: not how tightly a file can be packed, but whether the rule behind it can be found.

INEN is the fifth in a sequence of methods filed and deposited by the inventor since June 2026, each verified before the next was built on it. This release follows the project's initial disclosure on the KTR framework.

###

About the Knowledge Triangle Route

Created in the United Arab Emirates by sole inventor Roman Burtsev, KTR is a universal methodological framework designed to translate complex phenomena into computable quantities. Using three-state logic and structured algorithmic reduction, it provides the scientific and engineering communities with verifiable tools for storage, transmission, and computation.

Media Contact:

Leila Salieva

Email: press@ktr-project.org

Website: https://www.ktr-project.org