



# Algoland Compressor

Measured compression brief for files, folders, archives and enterprise storage validation

Public product	Algoland Compressor
Domain	<a href="https://algoland.us/">https://algoland.us/</a>
Artifact	.algd verified package
Delivery	sealed runtime, CLI and API
Claim discipline	measured rows only, SHA-256 verified restore

**19.44%**

Cubbit raw reduction

**104/104**

SHA-exact public sweep

**.algd**

artifact extension

This document is public-safe. It does not disclose source code, implementation internals, private formulas, or unverified universal wins. All compression figures below are counted-byte measurements with restore verification requirements.

# 1. Product Surface

Algoland Compressor is the current public product. It provides local CLI and API workflows for files, folders, archives, mixed payloads, .algd packaging, reporting and SHA-256 restore verification.

## Files and folders

Individual files, folders and mixed enterprise payloads can be processed through the same public runtime boundary.

## Archive-aware intake

ZIP, RAR, TAR, TAR.GZ, TGZ and archive bundles are accepted by the operational layer for packaging, accounting and restore validation.

## Verified artifacts

.algd artifacts are restored and checked against the original by SHA-256 before any win is treated as valid.

## Public endpoint examples

Endpoint	Purpose
GET /api/products	Public product metadata: one product, Algoland Compressor, .algd artifact extension.
GET /api/algoland/status	Runtime status surface for pilot validation.
POST /api/algoland/validate	Payload class and deployment validation request.
POST /api/algoland/quote	Estimate workflow for measured pilot planning.
POST /api/algoland/chunks/init	Chunked large-object workflow start.
POST /api/algoland/verify	Restore verification and SHA report.

The CLI is a thin public client for Algoland Compressor. It sends work to the local or sealed runtime, manages package workflows and restore verification, and does not expose internal compression logic.

## 2. Cubbit Validation

Across the three Cubbit validation datasets, the measured Algoland Compressor path reduced storage from 97,650,936 B to 78,670,263 B, a measured 19.44% reduction with SHA-exact restore and raw/no-gain protection.

Dataset	Raw / previous bytes	Measured Algoland bytes	Delta	Status
DS1	52,428,960 B	52,429,120 B	+160 B	Protected raw-floor case; no fake compression claim.
DS2	37,548,410 B raw	20,674,714 B	-16,873,696 B	Improved.
DS3	7,673,566 B raw	5,566,429 B	-2,107,137 B	Improved.
Aggregate	97,650,936 B	78,670,263 B	-18,980,673 B	19.44% raw reduction.

**97,650,936 B**

Cubbit raw total

**78,670,263 B**

measured output

**18,980,673 B**

bytes saved

The DS1 result is intentionally described as protection, not smaller compression. Already-incompressible or no-gain data must round-trip exactly and avoid false savings claims.

---

## 3. Public Sweep and Filetype Coverage

The representative public sweep covers 104 real files across 45 extensions, with SHA-exact restore on every row. The public-safe claim is  $\text{best\_of}(\text{Algoland}, \text{xz}, \text{raw}) \leq \text{xz}$  on that measured sweep; it is not a claim that the core alone beats every codec on every file.

Metric	Measured value
Files	104 real files
Extensions	45
Size range	1 B to approximately 2 MB in the public sweep; larger engineering tracks are maintained separately.
Restore	SHA-exact 104/104
Selector claim	$\text{best\_of}(\text{Algoland}, \text{xz}, \text{raw}) \leq \text{xz}$ on 104/104 measured rows
Core wins	80 rows
xz child wins	6 rows
raw-floor cases	18 rows

Covered classes include text, logs, CSV, JSON, XML, source/code, binary, object files, archives, images, audio, PDFs, office-like containers, generated data and already-compressed payloads. Coverage expands only when the rows are measured and restore-verified.

## 4. Active Absorption Research

The research line is loss-driven: every remaining loss is classified by mechanism, implemented as a general absorber, tested against the target rows, then rerun through no-regression gates. Public numbers change only after counted-byte, SHA-exact verification.

Stage	Measured result	Status
Stage 5A numeric delta	Closed g_tone.wav and g_weights_small.safetensors; SHA-exact and regression-gated.	Banked research win.
Stage 5C CSV column contour	g_sales_1m.csv: 81,831 -> 11,441 B; g_dump.sql: 3,939 -> 859 B.	Brotli open count reduced by 2.
Stage 5D token contour	g_vector_256k.svg: 30,393 -> 28,277 B; multiple HTML/JSON/XML/log rows closed.	xz open 7 -> 6; brotli open 30 -> 24.
Stage 5H ZIP/container contour	h_archive.zip: 1,768,717 -> 1,768,276 B; h_office.xlsx: 19,792 -> 19,773 B; g_document.odt: 4,950 -> 4,949 B.	Banked as 461 B gap reduction; no new row closures.
Current active gate	xz 6 open / brotli 24 open in the engineering absorption suite.	Still open; not a universal victory claim.

Next engineering targets: TIFF/image-plane prediction, parquet/container pages, ZIP/GZ/MP4/WebP archive handling, object-code/section modeling, and PAQ/ZPAQ/cmix-class context modeling.

# 5. Speed and Large Objects

The research runtime is not the final production architecture. Current benchmark harnesses often include process startup, full SHA roundtrip, competitor subprocesses and script overhead. The production target is a persistent, chunked, streaming runtime with parallel workers and tree verification.

Track	Measured state / goal
47-row speed matrix	0 missing Algoland byte rows; 0 xz open, 20 brotli open, 7 zstd open across generated code/text/structured/binary/audio/archive rows.
Raw-orbit R0	Banked for random/high-entropy/no-gain rows: random_1048576.bin went from 14.964 s to 0.054 s in the R0 gate, same raw+8 artifact.
Archive/container R1	Banked narrowly on archive_mixed.zip: 131,458 -> 131,366 B versus the previous Algoland speed path, and 4.784 s -> 1.507 s with SHA-exact restore.
Persistent runtime R2	Next speed target: avoid one process per file; run batches, folders and archive sessions inside a long-lived sealed runtime.
Chunked streaming	Content-defined or policy-defined chunks, parallel workers, resumable sessions and SHA tree verification.
Large objects	Support storage payloads that exceed local memory by streaming and manifesting .algd chunks.
Already-compressed input	Protect no-gain cases through raw/no-gain handling instead of reporting fake wins.

Speed claims are treated like compression claims: measured rows only. The current speed ladder has banked raw/no-gain recognition and the first mixed-ZIP container fast path. Broader archive families and the persistent runtime remain active engineering targets.

## 6. Deployment and IP Boundary

If a customer uses its own compute, executable code must run in the customer environment. The protection comes from the delivery model: sealed runtime, signed build, limited operational API, integrity checks and legal controls. Algoland does not promise that reverse engineering is impossible; it avoids disclosure of source and private internals.

Model	Where data runs	What customer receives
Remote API	Algoland infrastructure	HTTPS API access only; useful for synthetic tests or samples that can leave the customer environment.
Locked container or VM	Customer infrastructure	Signed sealed runtime exposing local API endpoints; balanced pilot model.
Confidential runtime	Customer cloud or dedicated host	Attested runtime with encrypted storage and restricted access where supported.
Algoland appliance	Inside customer network	Local network API with controlled runtime delivery and support.

Shared with customers: API endpoints, usage reports, validation reports, restore verification and deployment policy.  
Not shared: source code, implementation internals, debug access, compiler pipeline or private documentation.

# 7. Public Claim Boundaries

Allowed public statement	Boundary
Algoland Compressor supports files, folders, archives and .algd restore verification.	Operational product capability.
Cubbit validation aggregate improved 19.44% with SHA-exact restore.	Measured on the stated three datasets.
The public 104-file sweep is SHA-exact and best_of(Algoland, xz, raw) <= xz on all rows.	Measured selector claim, not universal core-alone dominance.
Stage 5H research gate is xz 6 open / brotli 24 open.	Engineering research status, not a sales guarantee.
Raw-orbit R0 improves random/no-gain speed.	Measured speed/runtime win, not a universal speed claim.
Archive/container R1 improves archive_mixed.zip.	Narrow measured ZIP-family speed/byte bank, not a claim that every archive format is solved.

Not claimed	Reason
Algoland beats every compressor on every file.	Not measured; some research gates remain open.
PAQ/ZPAQ/cmixon victory.	Those remain active research absorption targets.
Universal savings on all data.	Random, encrypted and already-compressed data may require raw/no-gain protection.
Source or internal formula disclosure.	Enterprise delivery uses sealed runtime and API boundaries.