

DETERMINISTIC RISK-COMPUTATION INFRASTRUCTURE · MAY 2026

ARIN22

Class-routed deterministic risk-compute kernel for institutional finance — with measured economic impact at the trading-desk and model-risk-management layer.

SAA Alliance builds **ARIN22** — a deterministic, class-routed risk-computation kernel validated end-to-end on an 8xH100 DGX-class node provided by the NVIDIA Innovation Lab. This brief sets out what the kernel does, what is closed pre-client, the canonical evidence behind it, and the institution-specific economic value a deployment produces at Tier-1 G-SIB scale — where capital efficiency and intraday latency dominate compute savings by 50–500x.

PRE-CLIENT · INSTITUTIONAL DILIGENCE-READY

1 What SAA Alliance does

SAA Alliance builds **ARIN22**, a deterministic risk-computation kernel. It produces tail-risk results — VaR, CVaR, FRTB-IMA tail-stress, intraday market risk, options — that the industry runs as overnight Monte Carlo. ARIN22 is class-routed: deterministic where deterministic provably publishes, CRN-anchored at the 99.9 tail, and routed to a 10M Monte Carlo challenger on hard regime-shift cases with every routing decision hash-logged. Every result is bit-exact reproducible and emitted under fail-closed governance — the kernel refuses to publish a decision-grade result rather than degrading silently. ARIN22 is positioned to be embedded inside or alongside Aladdin, Bloomberg, Murex and MSCI — not to replace them.

2 Differentiation

■ Class-routed, not single-method

Deterministic · CRN-anchored · 10M MC challenger · BLOCK — every routing decision hash-logged. No silent degradation.

■ Sub-millisecond p99 by construction

Tick replay p99 0.44–0.46 ms; pre-trade gate p99 0.85–0.93 ms with p999 1.8–2.5 ms (disclosed) on 60M-order fixture. Validated on canonical 10M MC.

■ CPU-portable kernel, no GPU lock-in

Risk kernel validates identically on commodity ARM CPU and on 8xH100 — difference is throughput, not the answer.

■ Audit-grade by construction

HMAC-signed verdict trail · ed25519 ledger · cryptographic hash-traceable lineage. Built for SR 26-2 / FRTB-IMA model-risk review.

3 Validation status

<p>8xH100</p> <p>80 GiB · DGX-class · Innovation Lab</p>	<p>11</p> <p>institutional asset classes validated</p>	<p>0.44–0.46 ms</p> <p>tick replay p99 · 24M updates</p>	<p>58 / 58</p> <p>data-room artifacts · 0 missing</p>
---	---	---	--

Enterprise Wave canonical evidence (portfolio surface) closed May 2026 — ~27,700 distinct portfolio cases · 8.8B paths/backend · CRN 99.9 GPU↔CPU parity PASS · CVaR 99.9 vs 10M MC max diff **0.035%** · 1 unique hash across 1,000 fresh-process repeats. Canon Upgrade multi-asset wave (30 May 2026) — 11 asset classes (equities · ETFs · fixed income · FX · commodities · energy · precious metals · sovereign · crypto · multi-asset · options) · 10M MC with **N=3 fresh-process repeats** · CVaR99.9 GPU↔CPU parity p95 ~**0.27%**. Shadow Ledger armed (1 / 250 matched obs). Full evidence pack at saa-alliance.com/arin22-demo.

4 Founder

The founder previously led national liquidity management at PrivatBank, Ukraine's largest bank — with responsibility for national-level physical cash reserves — and is a published academic researcher.

ARIN22 comes directly from operating institutional risk and liquidity at sovereign scale, not from adjacent tooling.

- NATIONAL LIQUIDITY MANAGEMENT
- SOVEREIGN-SCALE CASH OPERATIONS
- PUBLISHED RESEARCH · CFA

5 Same-surface benchmark — industry fast methods vs ARIN22 governed promotion

SUBJECT-STRESS SURFACE · SAME H100 EXECUTION · 10,080 ROWS · 28 SCENARIOS × 6 HORIZONS · 10M MC REFERENCE

METHOD	CVAR99 P95 ERROR VS MC	CVAR99 MAX ERROR
Cornish-Fisher · industry-standard fast method	73.62%	74.58%
Delta-Gamma proxy · industry-standard fast method	17.58%	19.12%
ARIN22 governed deterministic promotion	0.289%	< 1%

Same compute, same surface, side-by-side. The economic translation of this accuracy gap is in Section 6, bucket B.

6 Economic effect — institution-specific estimate at Tier-1 G-SIB scale

ECONOMIC VALUE · INSTITUTION-SPECIFIC ESTIMATE · TIER-1 G-SIB

Indicative annual economic value across five mechanisms at the scale of a single Tier-1 G-SIB (\$500B+ balance sheet, \$50B+ trading book, 30–50 person quant organisation). Each row is anchored to validated evidence + an explicit assumption. Calibrated per-customer numbers are produced during the 90-day co-validation.

MECHANISM	INDICATIVE RANGE	ANCHOR & ASSUMPTION
<p>A. Capital efficiency</p> <p>Faster CCAR / IMM / FRTB-IMA → RWA optimisation, tighter capital allocation, aggressive CVA / XVA hedging.</p>	\$50M – \$500M / yr	Tier-1 G-SIB with existing IMA approval; ARIN22 acts as deterministic challenger improving tail-estimation precision. CVaR 99.9 within 0.035% of 10M MC on Portfolio surface. Capital relief subject to regulatory acceptance.
<p>B. Trading-desk latency value</p> <p>Intraday VaR at tick p99 0.44–0.46 ms vs overnight ~3 hr legacy recompute. Hedge before repricing; reduce stranded risk in regime transitions.</p>	\$5M – \$50M / yr	\$50B+ trading book; intraday repricing capture during regime transitions vs overnight reset. 24M tick updates · canonical p999 0.64–0.70 ms across N=3 fresh-process repeats.
<p>C. Quant productivity</p> <p>Backtest cycle: days → minutes. 10× experimentation throughput per quant.</p>	\$5M – \$20M / yr	30–50 person quant org. Deterministic replay (1 hash / 1,000 fresh-process repeats · 0.0% diff) enables faster iteration and audit-clean parallel experimentation.
<p>D. Audit & regulatory readiness</p> <p>Deterministic reproducibility + immutable per-run audit collapses Fed / OCC / FDIC SR 26-2, FRTB-IMA, MiFID II and EU AI Act validation cost.</p>	\$1M – \$5M / yr	5 decision-integrity invariants + 7 math-to-narrative audit rules + HMAC verdict trail + ed25519 hash-chained ledger shorten reviewer-side validation under existing customer MRM workflow.
<p>E. Compute cost reduction</p> <p>Minor opex line. Most G-SIBs run amortised owned hardware (\$5–\$15 / hr per H100), not AWS on-demand pricing.</p>	\$0.2M – \$1M / yr	70% MC-tail replacement on amortised owned-hardware baseline. CPU-portable kernel means GPU footprint does not scale with the maths.
TOTAL ANNUAL VALUE · PER INSTITUTION	\$60M – \$575M	Aggregate across mechanisms A–E at Tier-1 G-SIB scale per assumptions above.

KEY MESSAGE

The dominant economic value of ARIN22 is **not** compute-cost reduction. Capital efficiency and trading-desk latency value dominate compute savings by **50–500×**.

Disclosure discipline. Ranges above are indicative for a Tier-1 G-SIB profile, anchored to validated evidence and explicit assumptions — they are not customer-calibrated forecasts and not a commitment. Production numbers per book are produced during the 90-day technical co-validation. Two surfaces operate at two tolerance budgets: Portfolio-surface CVaR within 0.035% of 10M MC; Subject-Stress surface governed promotion within 6.35% p95 / 9.79% max. The figures are not interchangeable.

7 NVIDIA stack alignment

The risk kernel is CPU-portable by design; where the architecture consumes NVIDIA is the inference and agentic layer. That footprint scales with client traffic as pilots convert, not with the kernel itself.

Compute	8×H100 (80 GiB) DGX-class node — NVIDIA Innovation Lab. Used for batch scale, validation runs, GPU↔CPU parity proofs.
Inference	NVIDIA NIM API across 5 of 7 production modules · NeMo Guardrails as the narrative-gate (anti-hallucination) layer across 2 of 7 · vLLM on H100 as the local-sovereignty primary path.
Data stack	RAPIDS / cuDF for ontology pre-processing on the typed-relation knowledge graph that grounds the agentic council.
Programs	NVIDIA Inception member · Innovation Lab participant with the H100 DGX-class node above as the canonical validation substrate.

8 What we are asking

First pilot · 90-day technical co-validation

A 90-day technical co-validation engagement on a real institutional book — Tier-1 G-SIB / SWF / reinsurer / central-bank model-risk. Air-gapped or in-tenant deployment supported day one. Produces customer-calibrated economic numbers vs the indicative ranges in Section 6.

Peer-grade technical introduction

An introduction into the institutional risk or model-risk-management function of an appropriate Tier-1 institution. A peer-grade technical conversation, not a sales call. Live walkthrough of the data room available on request.

9 What is not claimed

No external certification

No STAC certification; Heston / options work is STAC-A2-inspired, self-run on 8×H100. No NVIDIA-attested benchmark.

No realized regulatory backtest pass

Kupiec · Christoffersen · Acerbi-Szekely structurally gated until 250 matched observations. Shadow Ledger armed; currently 1 / 250.

No live-broker production deployment

Status is PRE-CLIENT. Latency figures are internal evidence on a controlled test substrate, not live execution telemetry.

No bank-client production sign-off

Customer-side model-risk-management sign-off is the deliberate next step. First pilot ask above is exactly that conversation.

10 Positioning note

SAA is risk *infrastructure*, not an AI platform. ARIN22 is a deterministic mathematical kernel; LLM-based reasoning is layered on top only as a constrained, audited narrative — never as decision authority. The kernel is CPU-portable by design: *GPU is for scale, not for kernel correctness* — no GPU lock-in. ARIN22 is positioned to be embedded inside or alongside Aladdin, Bloomberg, Murex or MSCI, not to replace them.

PUBLIC REFERENCES

Data room · Layer 1: saa-alliance.com/arin22-demo — Institutional Evidence Pack, six chapters, canonical numbers and validation surfaces.

Use-case map: saa-alliance.com/use-cases — deployment patterns by institution type and risk surface.

MATH FIRST · AGENTS SECOND · GOVERNOR ALWAYS · AUDIT FOREVER