Further Google RE2 Benchmark Results Prove Dramatic Infrastructure Cost Savings

A single Nol8 FPGA appliance has been benchmarked to replace the equivalent compute of up to 60,000 CPU's

Highlights

- Further benchmark testing has demonstrated that a **single Nol8 FPGA appliance** has been benchmarked to replace the equivalent compute capacity of **up to 60,000 CPU's** under AI-grade workload conditions (P99 load at 6,000+ rule complexity).
- **Nol8's world-first technology has unlocked a previous ceiling to infrastructure scalability** delivering AI-grade data processing at a fraction of the cost and footprint of conventional CPU-based infrastructure solutions.
- A conventional 10,000 CPU array for high-complexity AI workloads carries an estimated annual operating cost of **A$4.5m**. Based on the further benchmark testing, a single Nol8 FPGA appliance delivers equivalent capacity for **under A$37,000 per year in direct hardware costs.**
- Nol8's AI Data Plane is purpose-built to solve the data processing constraints of unstructured AI workloads which is building rapidly to represent **90% of all future data flows.**
- Results published represent only a portion of Nol8's total performance capability — future benchmark testing will be conducted and published as the Company continues to explore the full limits of its world-first technology.

FortifAI Limited (ASX: FTI) ("**FTI**" or the "**Company**") is pleased to provide an update on the infrastructure economics results from its benchmarking of the Nol8 AI Data Plane against Google's RE2 engine, a widely accepted industry standard framework for data pattern-matching.

Further to the Company's previous announcements, the results of the Company's most recent infrastructure testing confirms that a single Nol8 FPGA appliance can replace the equivalent compute capacity of up to 60,000 CPU's under the most demanding real-world AI data conditions. The infrastructure focused results represent a world-first technological breakthrough in the economics of AI data infrastructure.

CPU Replacement Results: One FPGA vs traditional CPU Arrays

The following table shows the number of CPU's required to match the throughput of a single Nol8 FPGA appliance at each complexity tier and load condition.

COMPLEXITY TIER	P50 (CPU's Required)	P90 (CPU's Required)	P99 (CPU's Required)	NOL8 FPGA APPLIANCE	P99 RESULT
Low (~10 rules) Basic Web / API Applications	30	57	116	**1 FPGA**	**Replaces 100 CPU's**
Medium (~1,000 rules) Enterprise security	824	2,063	8,251	**1 FPGA**	**Replaces 8,200 CPU's**
High (6,000+ rules) AI-grade data classification	1,124	3,555	60,359	**1 FPGA**	**Replaces 60,000+ CPU's**

Source: Company commissioned RE2 benchmarking. CPU equivalency calculated from throughput parity at each percentile and complexity tier. Figures are estimates.

"These results reframe how enterprises should think about AI infrastructure investment. The question is no longer around how many CPU's do we need? It is about why are we still using CPU's at all for this class of workload. A single FPGA appliance replacing 60,000 CPU's is not an incremental efficiency gain. It is a structural shift in the economics of data infrastructure.

Every CPU freed by Nol8 is a CPU that is no longer bottlenecking a GPU. Enterprises have invested heavily in GPU capacity to run their AI workloads, but when the data pipeline feeding that GPU is CPU-bound, the GPU is not being pushed to its potential. Nol8 solves the bottleneck that is preventing the AI investment enterprises have already made from delivering its full return"

Dr. Alon Rashelbach, Co-Founder and CTO, Nol8

The Infrastructure Economics: 10,000 CPU's vs 1 FPGA Appliance

To contextualize the benchmark results commercially, consider a mid-scale enterprise deployment of 10,000 CPU's processing high-complexity AI data workloads. This is a conservative example as many of the use cases noted below require significantly greater infrastructure.

COST ITEM	10,000 CPU ARRAY	NOL8 SINGLE FPGA APPLIANCE
Hardware OpEx (cloud computing)	~A$4.5M/yr cloud compute	~A$37,000 per year
Infrastructure Management	A$1.5M–$3M+ per year	<A$1,000
Estimated Total Annual OpEx	A$6-7.5m+ per year	< A$50,000 per year

Source: Amazon Web Services Publicly Available pricing as at 27th April 2026. Cost estimates based on AWS cloud compute pricing (F2 and C6 instance families).

Where Does a 60,000 CPU Arrays Occur in the Real World?

The 6,000-rule, P99 scenario directly reflects the workloads of modern AI-driven enterprises across multiple sectors. A small handful sample of real world scenarios are detailed below:

- **Cybersecurity and SIEM platforms** — Platforms such as IBM QRadar, Microsoft Sentinel, and Palo Alto Cortex continuously process thousands of detection rules across petabytes of log and event data. Large enterprise deployments routinely operate arrays of thousands to tens of thousands of CPU's for this function alone.
- **Financial services fraud and compliance monitoring** — Real-time transaction screening across global payment networks requires simultaneous matching against thousands of regulatory, fraud, and sanctions rules. Tier 1 banks operate CPU arrays of equivalent scale for this function.
- **Telecommunications deep packet inspection** — Carriers performing network-level inspection at 100 Gbps+ for lawful intercept, quality-of-service enforcement, and security deploy CPU infrastructure at equivalent scale.
- **Enterprise AI data pipelines** — As organisations deploy agentic AI and LLM systems, data flowing into and between models must be classified, filtered, governed, and routed in real time. This is precisely where Nol8's AI Data Plane operates — and where 6,000+ rule complexity is the standard.

About Nol8's Technology

Nol8 is creating a new category in enterprise technology: the AI Data Plane — the high-speed infrastructure layer that sits between raw data and AI inference, purpose-built to solve the data processing constraints of unstructured AI workloads. Built on algorithmically enhanced Longest Prefix Matching, scaled by machine learning neural networks, and hyper-accelerated by FPGA hardware, Nol8's engine processes data-in-flight at millisecond-grade latency

without buffering or batching. Backed by five years of published academic research from Technion University, Nol8 has unlocked a previous ceiling to data processing scalability that no software-based architecture has been able to surpass.

Unlike CPU-based software approaches, Nol8's FPGA architecture processes data in parallel at the hardware level. Performance does not degrade as workload complexity increases.

The AI Data Opportunity

The global datasphere is forecast to grow from 334 Zeta Bytes in 2025 to 19,267 Zeta Bytes by 2035, driven overwhelmingly by unstructured data generated by AI agents, autonomous systems, and large language models. Nol8's AI Data Plane is purpose-built to solve the data processing constraints of unstructured AI workloads — which will represent 90% of all future data flows.

This data must be filtered, enriched, classified, and routed in real time — before it reaches the model. This is the AI Data Plane: the missing infrastructure layer the industry now requires. Nol8's benchmark results confirm it has built this layer, and that its performance is in a category of its own.

Authorised for release by the board of FortifAI Limited.

Media and Public Relations contact:

Emily Walkerden (New York)
emily@nol8.io
+1 (917) 545 5817

About FortifAI

FortifAI Ltd (ASX: FTI) is an ASX listed AI infrastructure company developing and commercialising technology with a focus on AI. In addition to developing AI-forward technologies, FortifAI has developed a broad portfolio of video games. FortifAI uses AI to target efficiencies and expansion opportunities in technology.

Glossary

The following company and industry terms are used in this announcement:

Agentic AI: AI systems that operate autonomously and continuously, executing decisions and actions in real time without human input at each step.
AWS Bedrock Guardrails: Amazon's managed safeguard layer for generative AI applications, which uses RE2-based pattern matching to filter and govern data flows into and out of foundation models.
CPU's: Central Processing Units — the primary general-purpose processors used in conventional computing infrastructure.
FPGA: Field-Programmable Gate Array — a specialised semiconductor device whose hardware logic can be configured after manufacture, enabling parallel, high-speed data processing that far exceeds the throughput of conventional CPU's.
LLM: Large Language Model — an AI model trained on large volumes of data to understand and generate human language, such as OpenAI's GPT, Anthropic's Claude, and Google's Gemini.
P50: The median performance result — representing system behaviour under normal, typical operating conditions.
P90: The 90th percentile performance result — representing system behaviour under elevated, high-load operating conditions experienced in the top 10% of traffic events.
P99: The 99th percentile performance result — representing system behaviour under extreme load, reflecting the most demanding 1% of real-world operating conditions.
RE2 or RE2-based software: Google's open-source regular expression (regex) engine — an globally industry standard benchmark for data pattern-matching performance.
SIEM platform: Security Information and Event Management platform — enterprise software that aggregates and analyses security event data in real time for threat detection and compliance.
ZB: Zettabyte — a unit of digital data equal to one trillion gigabytes, used to measure global data volumes at scale