INFINIDAT

The Standard in Enterprise Storage

The Role of Storage in AI - Applications and Workloads

Eric Herzog
CMO
eherzog@infinidat.com
@zoginstor

FMS 2025
Session AIML-102-1
August 2025

Gen AI - Risk and Reward!

Bad or inaccurate responses leads to:

- Legal and Compliance Issues
- Brand Reputation/Perception Issue
- Customer Satisfaction
- Loss of business to Competitors
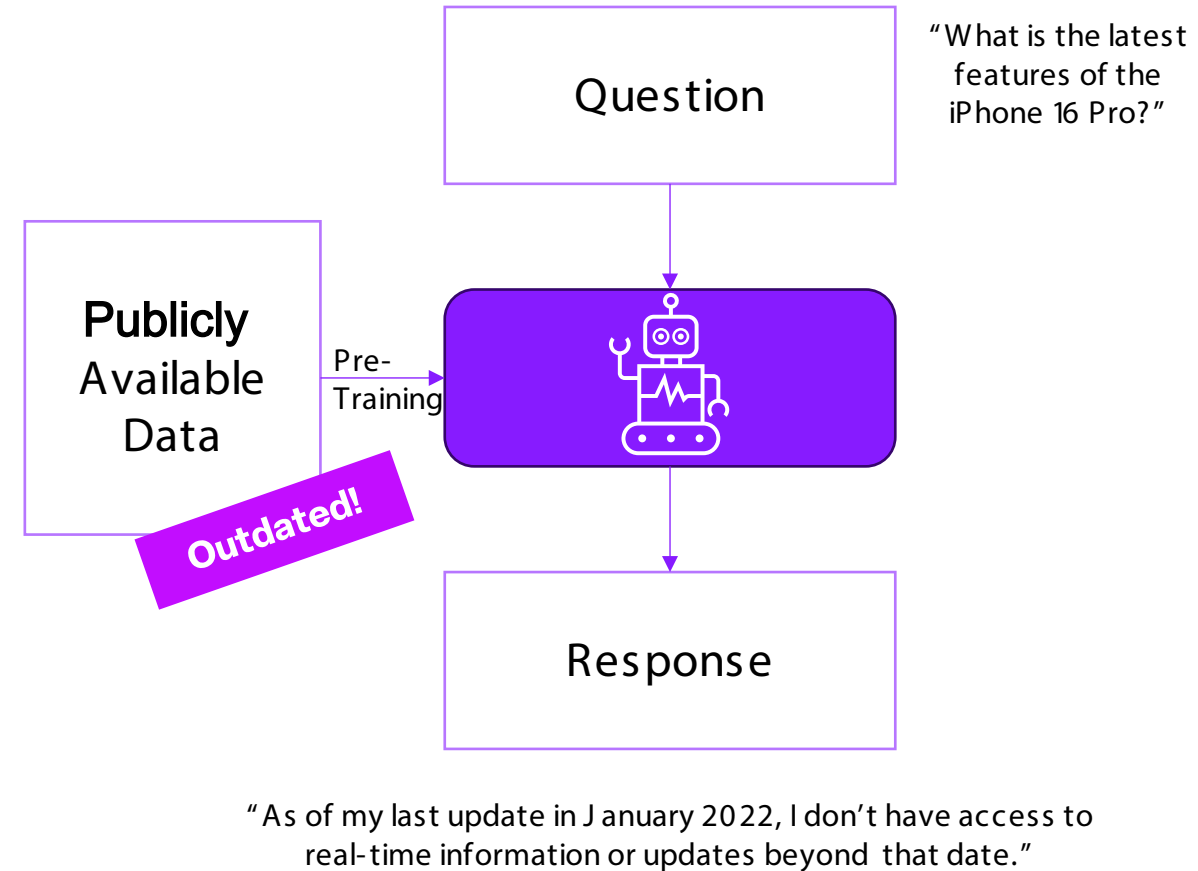
Accurate data produces

- Better Customer Experience
- Enhanced Productivity/Efficiency
- Higher Customer Satisfaction
- Innovate and Create Faster
- Cost Reductions and Higher Profits

"Amazon's (AMZN) AI upgrade of its Alexa device is expected to be delayed because of one rather inconvenient problem – it gives wrong answers to questions."
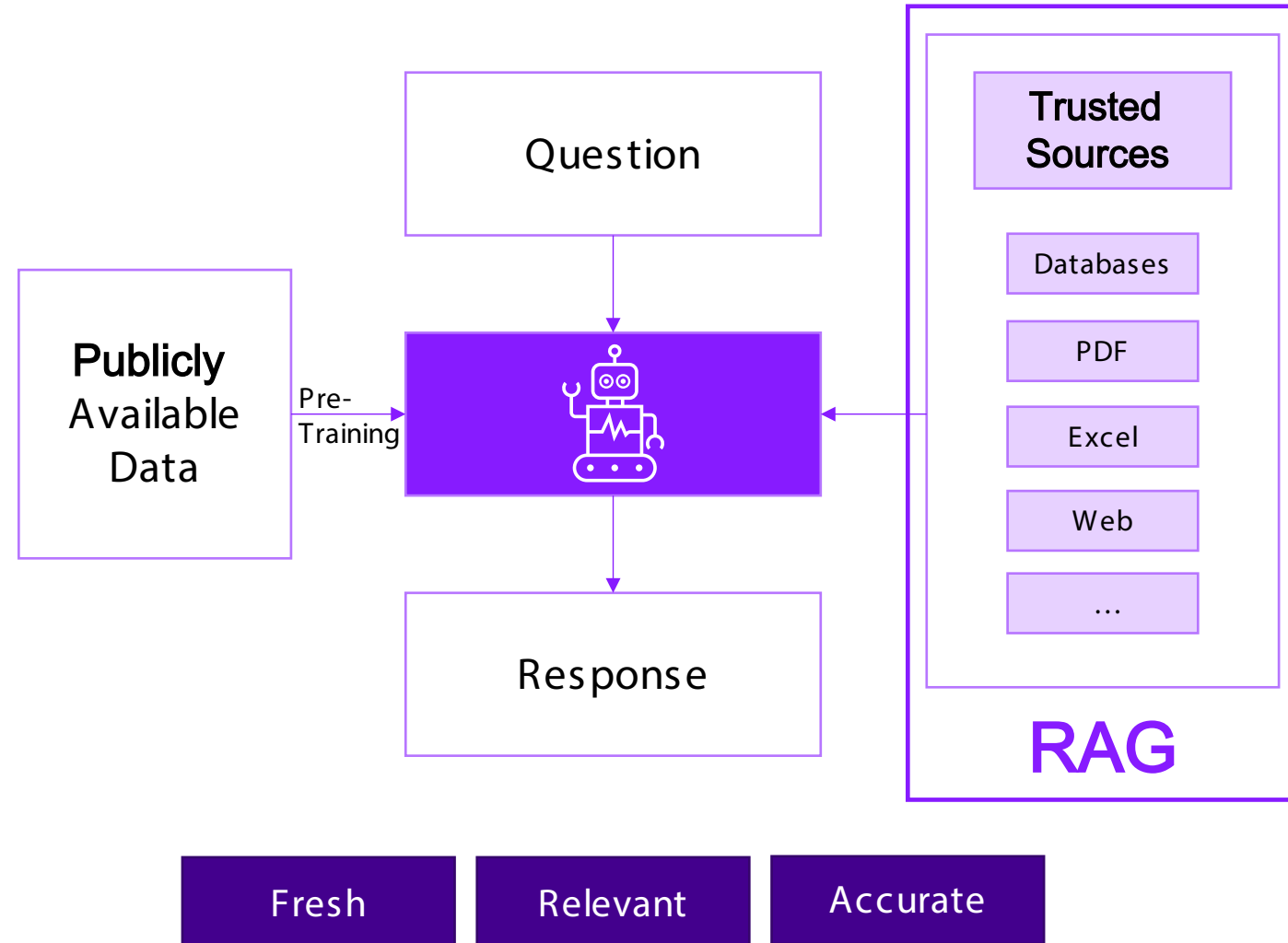
Gen AI – Getting the right answer is a challenge!

- LLMs and SLMs typically trained on publicly available data, lack exposure to an enterprise's specific and proprietary data

- Result - inaccurate and unreliable answers called "hallucinations" appear to be plausible answers but are incorrect or outdated

- Leads to misinformed business decisions, providing customers with wrong information which can damage the enterprise's reputation

- Gen AI needs continuous augmentation with new, trusted, and accurate data to enhance and maintain its reliability

Question

"What is the latest features of the iPhone 16 Pro?"

Publicly Available Data

Pre-Training

Outdated!

Response

"As of my last update in January 2022, I don't have access to real-time information or updates beyond that date."

R(etrival)A(ugmented)G(eneration) to the Rescue!

- Retrieval-Augmented Generation (RAG) combines the power of GenAI with an enterprise's private data to unlock fresh insights and streamline workflows

- Integrating information from external sources into LLMs and SLMs can achieve more accurate and specific results

Question

Publicly Available Data

Pre-Training

Response

Trusted Sources
Databases
PDF
Excel
Web
...

RAG

Fresh   Relevant   Accurate

Gen AI Meets Storage – More Challenges!

- AI initiatives are hitting a wall at the storage layer in a numerous areas

- Massive amount of checkpoint writes will overrun the storage array

- Expensive GPUs are waiting for data from Bogged Down storage systems.

- Huge power usage requirements for AI systems

- LLMs are susceptible to cyber attack

Microsoft deal propels Three Mile Island restart, with key permits still needed

By Reuters
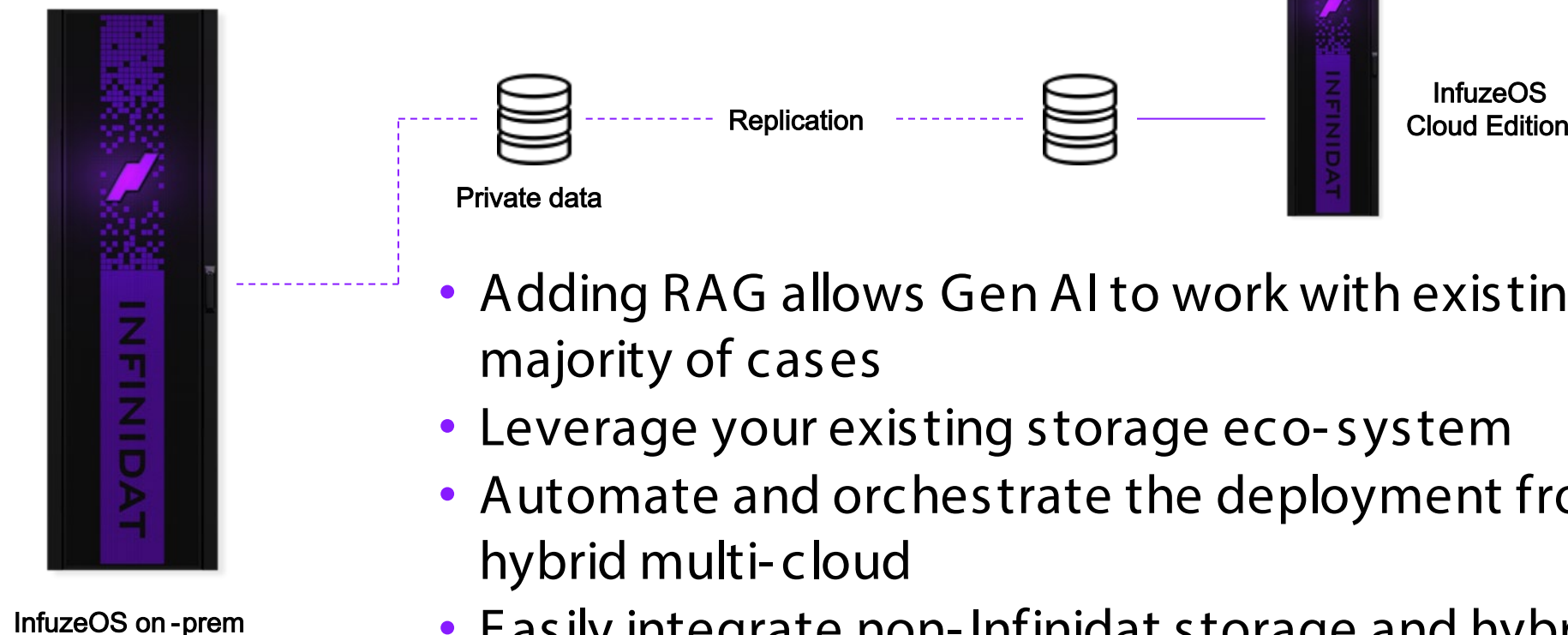
September 21, 2024 11:58 AM CDT · Updated 5 months ago

ARTIFICIAL INTELLIGENCE

AI is poised to drive 160% increase in data center power demand

Decoding the Double-Edged Sword: The Role of LLM in Cybersecurity

Infinidat Gen AI Solution

On-Premise

Replication

Private data

InfuzeOS
Cloud Edition

InfuzeOS on-prem
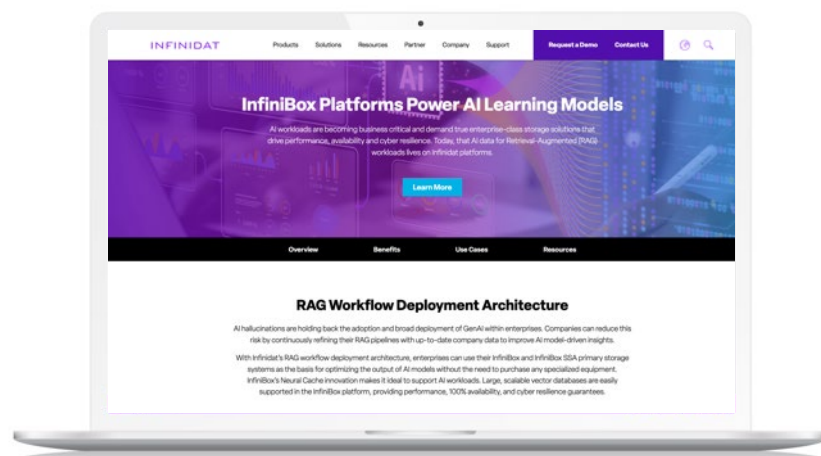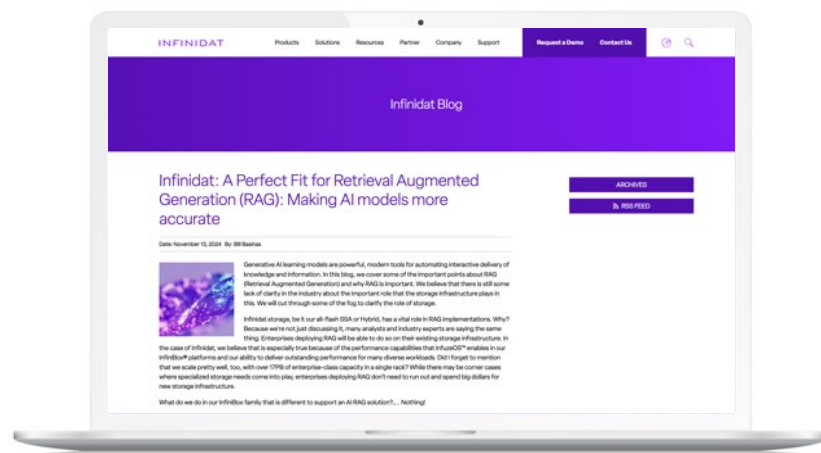
- Adding RAG allows Gen AI to work with existing infrastructure in the majority of cases
- Leverage your existing storage eco-system
- Automate and orchestrate the deployment from data center to hybrid multi-cloud
- Easily integrate non-Infinidat storage and hybrid multi-cloud environments that support NFS
- Agentic RAG adds new dimensions of flexibility and scalability

AI Solutions Resources

Blog :

https://www.infinidat.com/en/blog/infinidat-perfect-fit-retrieval-augmented-generation-rag-making-ai-models-more-accurate

Demo Videos:

https://info.infinidat.com/RAG_with_Infinidat_Reference_Architecture_Demo-EN_VideoLP.html

https://info.infinidat.com/Building_an_Agentic_AI_Workflow-EN_VideoLP.html

Analyst Whitepaper   :

https://www.infinidat.com/en/resource-pdfs/infinidat-cures-rag-response-time.pdf

Analyst Video :

https://blocksandfiles.com/2025/02/13/how-to-reduce-ai-hallucinations/

Solution Webpage :

https://www.infinidat.com/en/artificial-intelligence

Industry Acclaim: 2025

Seven Time Gartner® Peer Insights Customers' Choice Award Overall- 5 times

Finalist for Top Data Storage Innovation- InfiniBox® G4 family

2024 Si... InfiniBox... Flas...
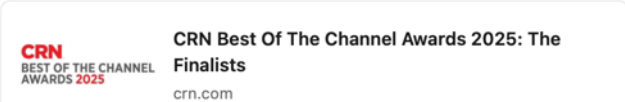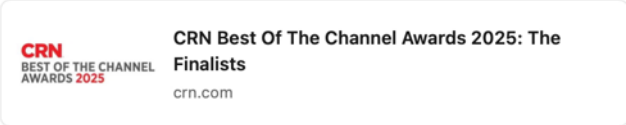
InfiniBox® G4 Family One of the 2025 TOP 5 Cybersecure Sub-10... Solu...

5-Star Rating in CRN® Partner Program Guide

3 Cyber Defense Magazine Global InfoSec Awards 2025:
- Hot Company for Cyber Resilience
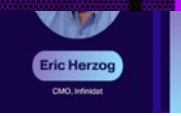- Editors Choice for Ransomware Recovery

INFINIDAT FINALISTS Storage Enterprise & Storage Cloud

CRN 2025 Tech Innovator Awards

CRN Best Of The Channel Awards 2025: The Finalists
crn.com

CRN Best Of The Channel Awards 2025: The Finalists
crn.com

Eric Herzog- 2025 CRN® Global Channel Chief

A Leader and a Fast Mover - 2025 GigaOm Sonar Report for Storage as a Service (STaaS)

OnCon 2025 Global Top 10 CMO/Marketers

CRN Best Of The Channel Awards 2025: The Finalists
crn.com
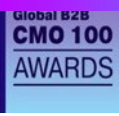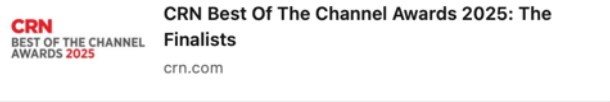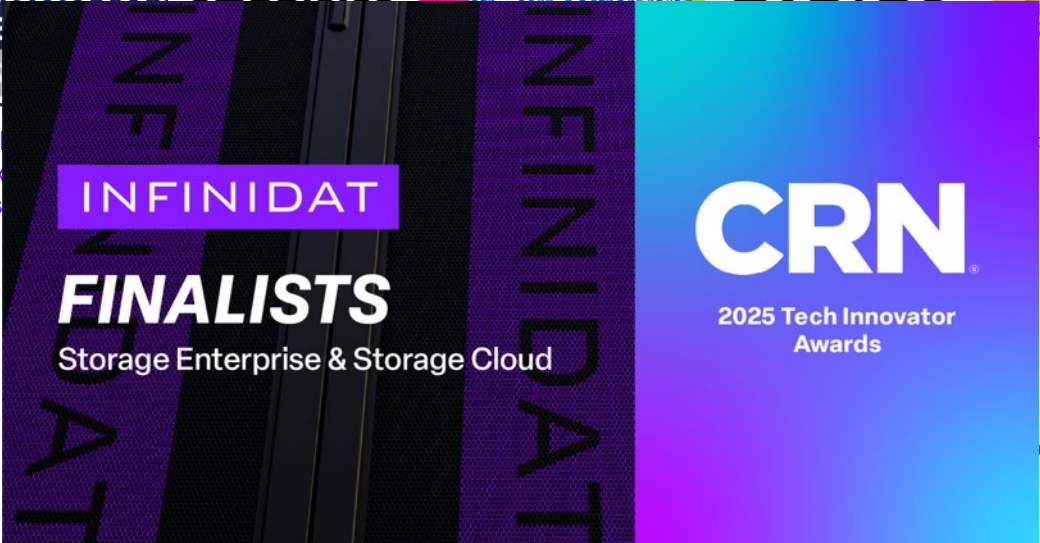
...nalist for Top Data ...otection Innovation - ...Safe® Cyber Detection

Eric Herzog- 2025 CRN® Inclusive Channel Leader

Finalist in 22 categories for Storage UK 2025 Awards including: InfiniBox G4, InfiniSafe and InfuzeOS Cloud Edition

Finalist for one of the World's Top 100 Marketing Teams

The End Users Have Spoken

InfiniBox Reviews

INFINIDAT

" InfiniBox SSA is an outstanding storage array."

Account Executive, IT Services | InfiniBox™ SSA

by Infinidat in Data Center - Others

5.0 ★★★★★ 21 Ratings

Gartner Peer Insights Customers' Choice 2025™

Gartner Peer Insights Customers' Choice 2024 North America™

Gartner Peer Insights Customers' Choice 2023™

As of July 28, 2025

Thank you

INFINIDAT