Hyperion Research:
ISC26 Market Update

June 2026

www.HyperionResearch.com
www.hpcuserforum.com

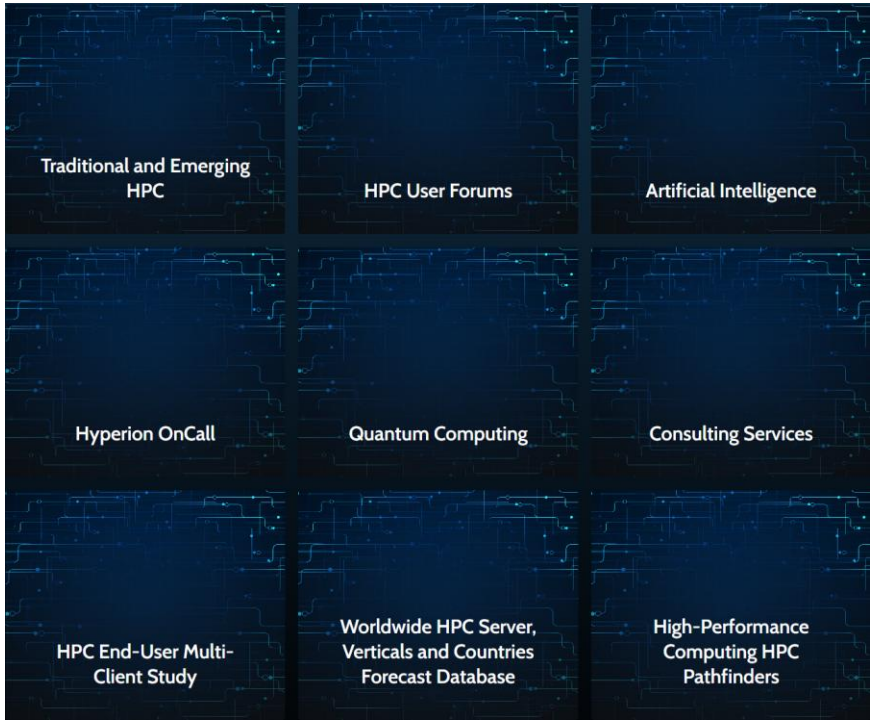
Earl Joseph, Bob Sorensen, Mark Nossokoff,
Tom Sorensen, and Jaclyn Ludema

Welcome

- Today's presentation will be sent to all registrants
- For follow-up details or discussions, please email mthorp@hyperionres.com
- The presentation will conclude at 8:30, but you are welcome to stay until 10:00 for networking and more coffee!

Who is Hyperion Research?

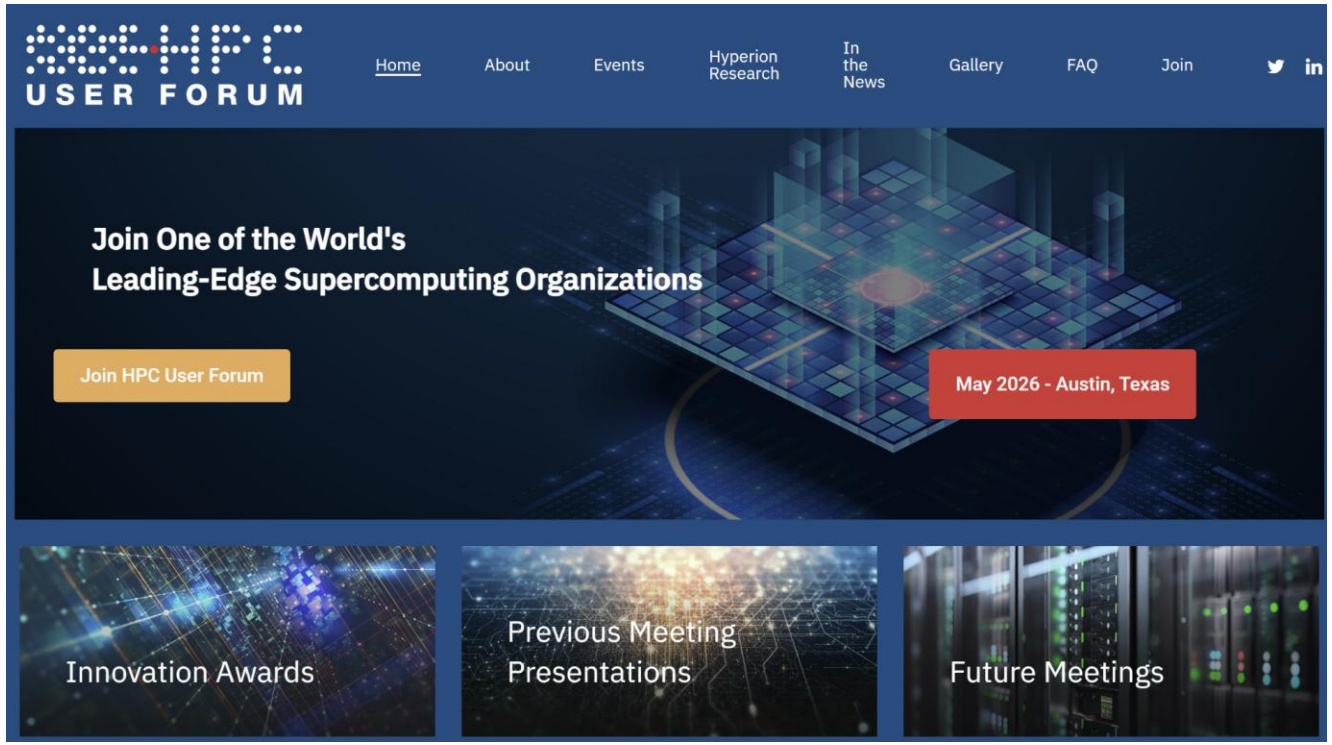
(www.HyperionResearch.com & www.HPCUserForum.com)

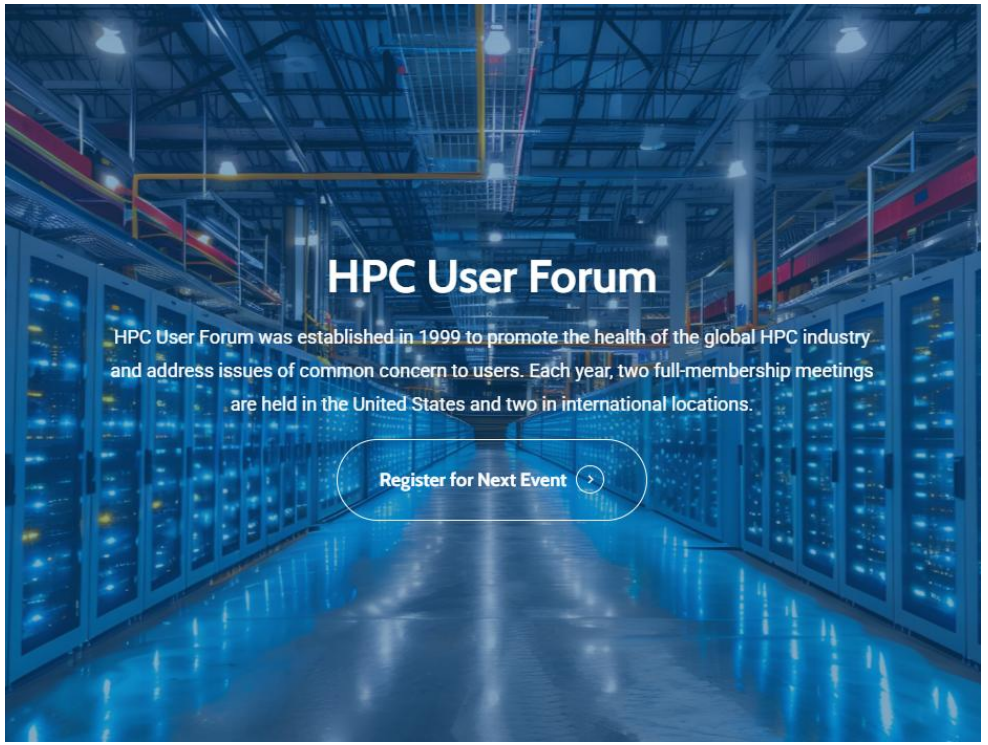


- **Provides timely, in-depth mission-critical insights across broad portfolio of relevant topics including HPC, AI, Quantum, storage and interconnects, data centers assessment, and ROI**
- **Delivered through variety of services consisting of custom projects, advisory, white papers, webinars, panel participation, and subscription**

Who is Hyperion Research?

(www.HyperionResearch.com & www.HPCUserForum.com)





- **Promote the health of the global HPC industry**
- **Address issues of common concern**
- **Hold multiple membership meetings throughout the year; typically, 2 in the U.S. and 2 globally**

Hyperion Research Team

Earl Joseph

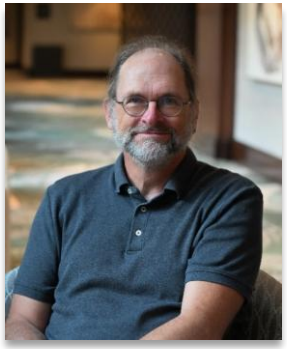
- CEO
- Executive Director HPC/AI User Forum

Jean Sorensen

- COO
- Human Resources

Bob Sorensen

- Sr. VP of Research
- Chief Quantum Analyst
- AI Analyst



Mark Nossokoff

- Research Director
- Chief Storage and Cloud Analyst

Jaclyn Ludema

- Sustainability Analyst
- Cloud Analyst

Tom Sorensen

- Principal AI/HPC Analyst
- Editor ACN Update

Mike Thorp

- Sr. Global Account Executive

Tyler Pla

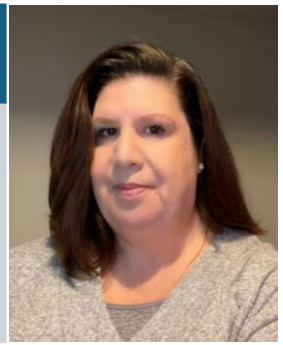
- Global Account Executive

Hyperion Research Partners

Katsuya Nishi

- Sr. Account Rep., Japan
- Chief Editor, HPCWire Japan

Mara Jacob



- HPC User Forum Event Director

Benjamin Portman

- President, Orpheus
- Web designer & developer

Andrew Rugg

- President, Certus Insights

Mike Heroux

- Consulting Analyst and Contributor

Kirsten Chapman

- Owner, K Chapman Consulting
- Data Collection Operations

KC Associates

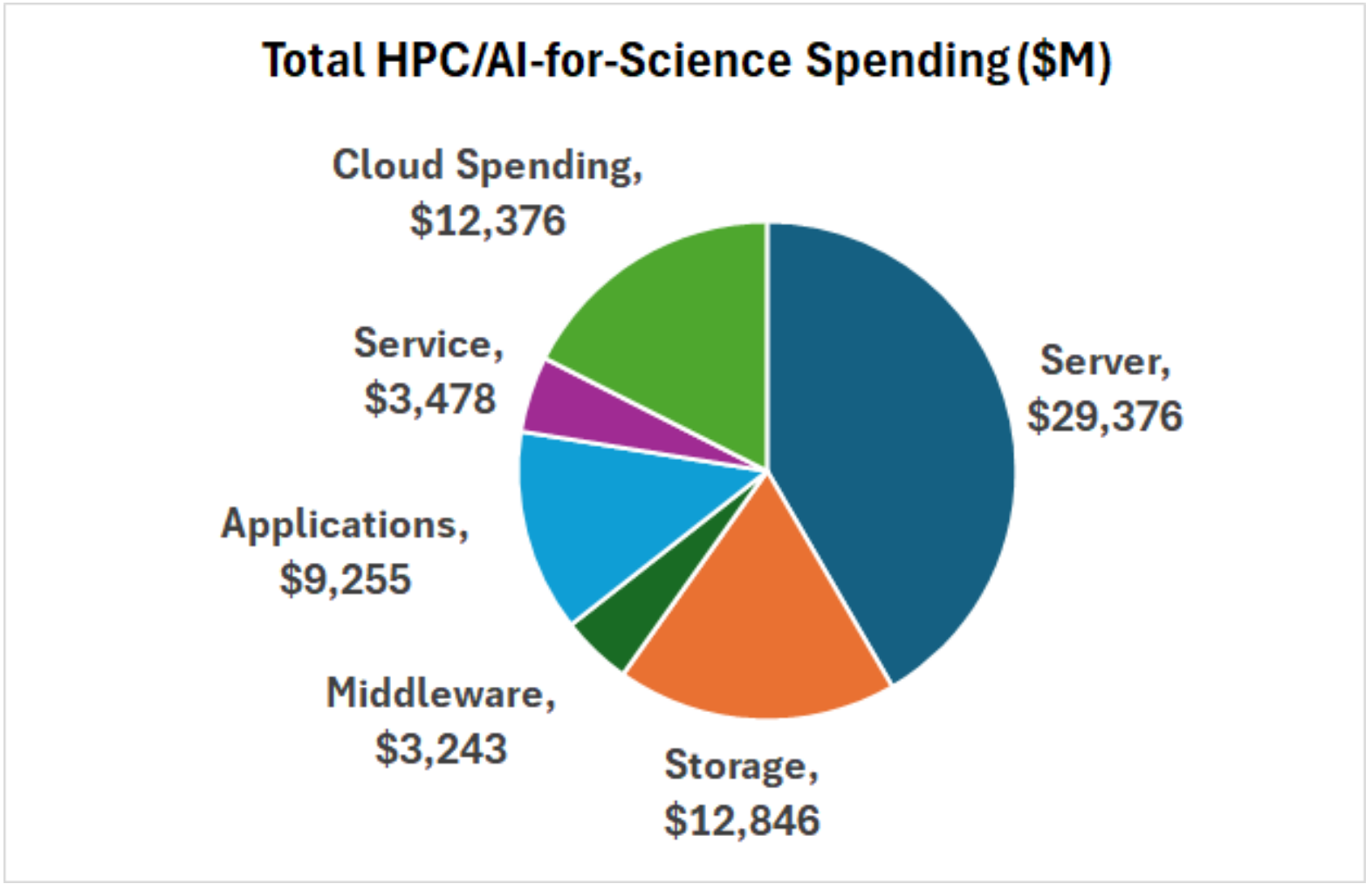
Today's Agenda

- **Mike Thorp, Senior Global Account Executive**
  - Introduction
- **Earl Joseph, CEO**
  - HPC and AI Market Update
- **Bob Sorensen, SVP, Chief QC & AI Analyst**
  - 6th Annual Global QC Market Survey
  - Recent Study Overviews: Gen AI ROI, AI in the Cloud
  - FP64 vs FP4: An Evolving Debate
- **Mark Nossokoff, Research Director, Chief Storage & Cloud Analyst**
  - Perspectives on HPC-AI Cloud, Storage, Interconnects, and Sustainability
  - Scientific Computing Cost, Value, and ROI Model
- **Conclusions**
- **Q&A**

2025 Was a Strong Growth Year

16.9% overall growth in spending!

- **15.0% growth in on-premises servers**
- **29.7% growth in the use of clouds**
- **Over $70 billion in total spending**



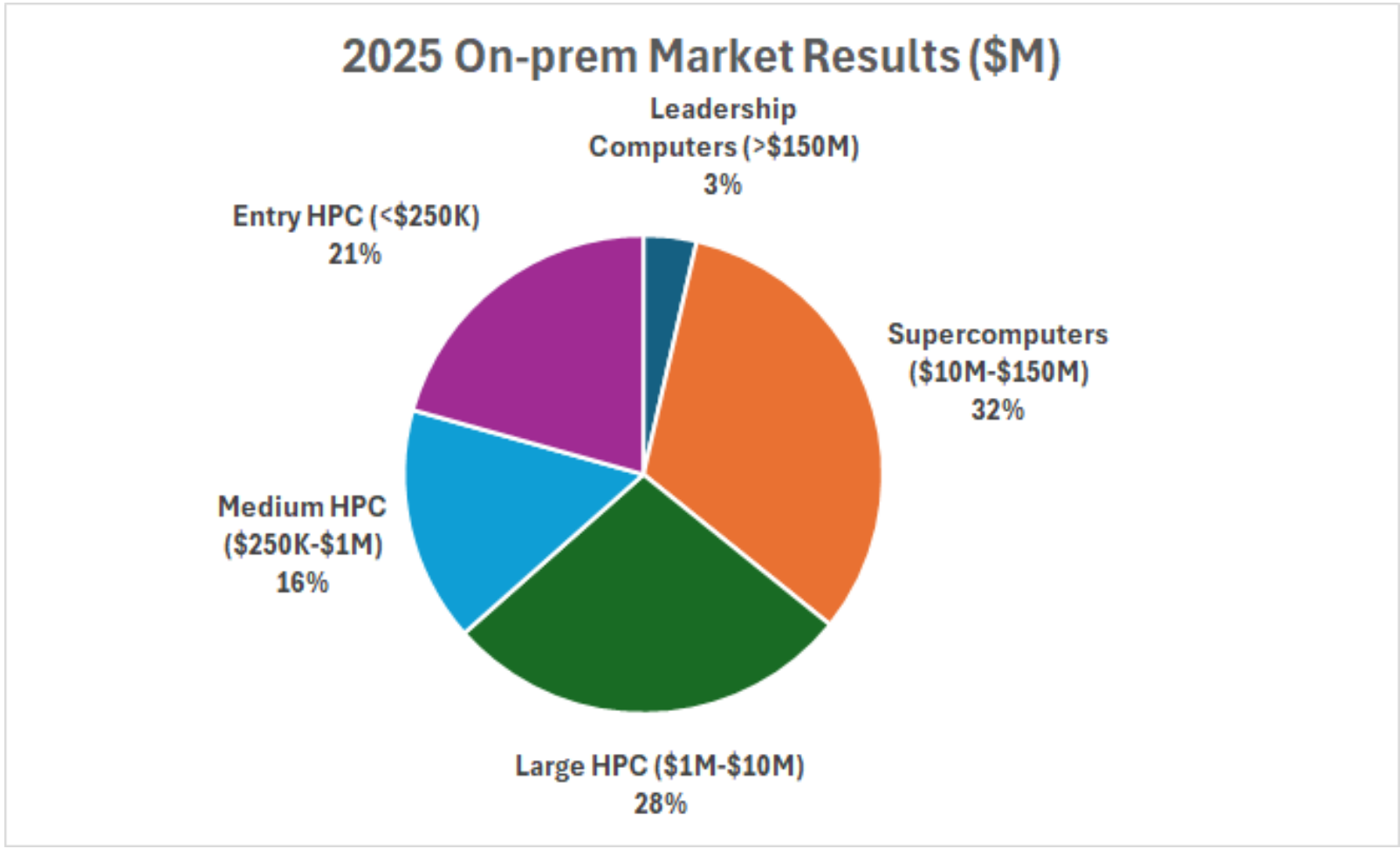
2025 HPC/AI Market By Vendor

On-prem sever sales were just under $30 billion in 2025

2025 HPC/AI On-Prem Sales		
Vendor	**2025 Sales**	**2025 Share**
HPE	$6,485	22.1%
Dell Technologies	$5,342	18.2%
Lenovo	$1,783	6.1%
Inspur	$1,236	4.2%
Sugon	$693	2.4%
Atos	$464	1.6%
IBM	$453	1.5%
Penguin	$508	1.7%
Fujitsu	$259	0.9%
NEC	$238	0.8%
Other HPC Suppliers	$2,735	9.3%
Non-Traditional Suppliers	$9,180	31.2%
Total	**$29,376**	**100.0%**
Source: Hyperion Research, April 2026		

2025 HPC/AI Market By Segment

Supercomputers ($10M-$150M) is the largest segment at $9.5 billion



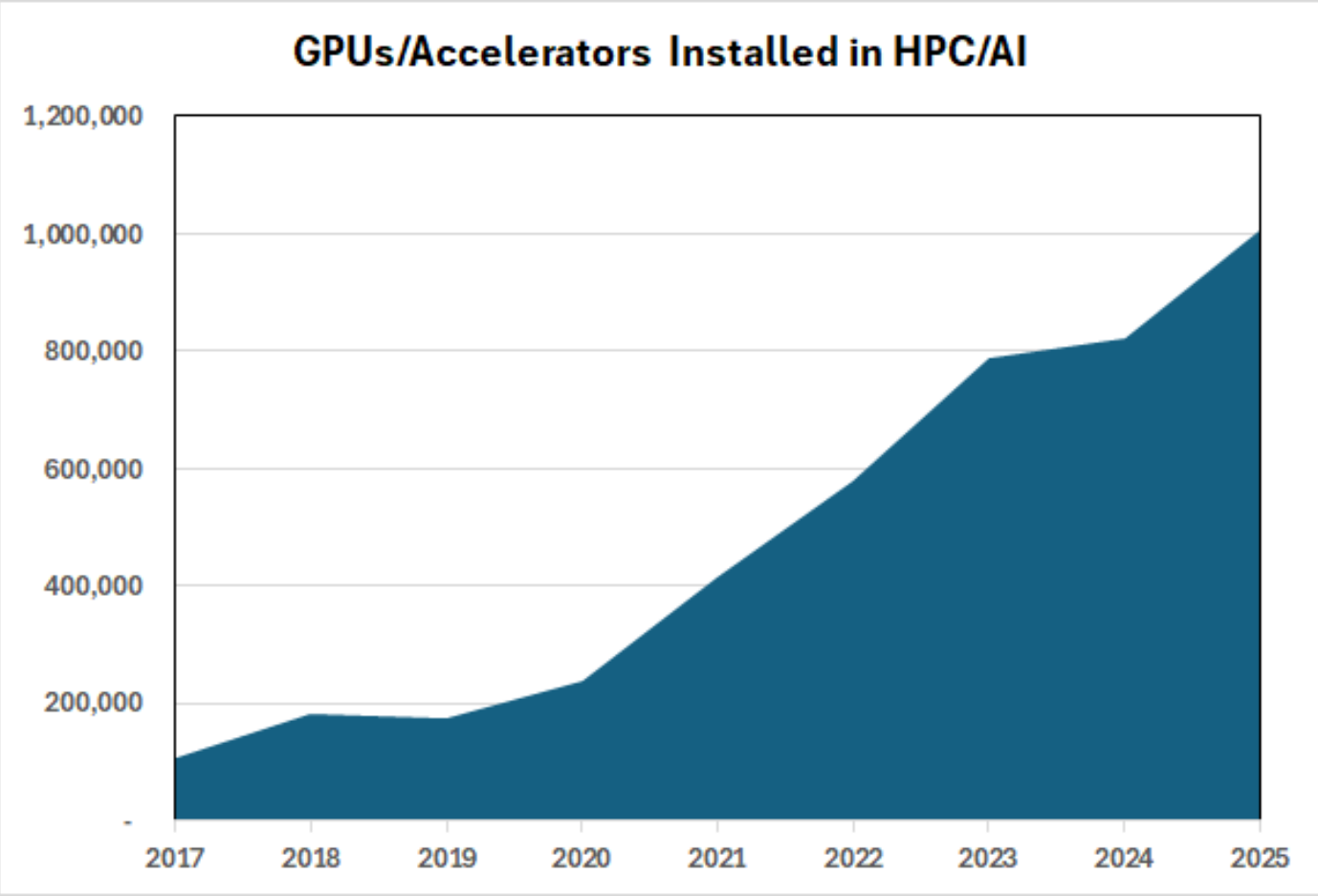
2025 HPC/AI Market By Vertical

Five verticals are now over $2 billion, and two are over $4 billion

HPC/AI On-prem System Installations ($M)		
	2024	2025
Bio-Sciences	2,302	2,622
CAE	2,751	3,134
Chemical Engineering	305	341
DCC & Distribution	1,384	1,610
Economics/Financial	1,336	1,526
EDA / IT / ISV	1,497	1,701
Geosciences	1,550	1,782
Mechanical Design	28	27
Defense	2,695	3,213
Government Lab	6,115	7,129
University/Academic	4,039	4,542
Weather	1,137	1,288
Other	415	462
Total Revenue	**25,554**	**29,376**
Source: Hyperion Research, May 2026		

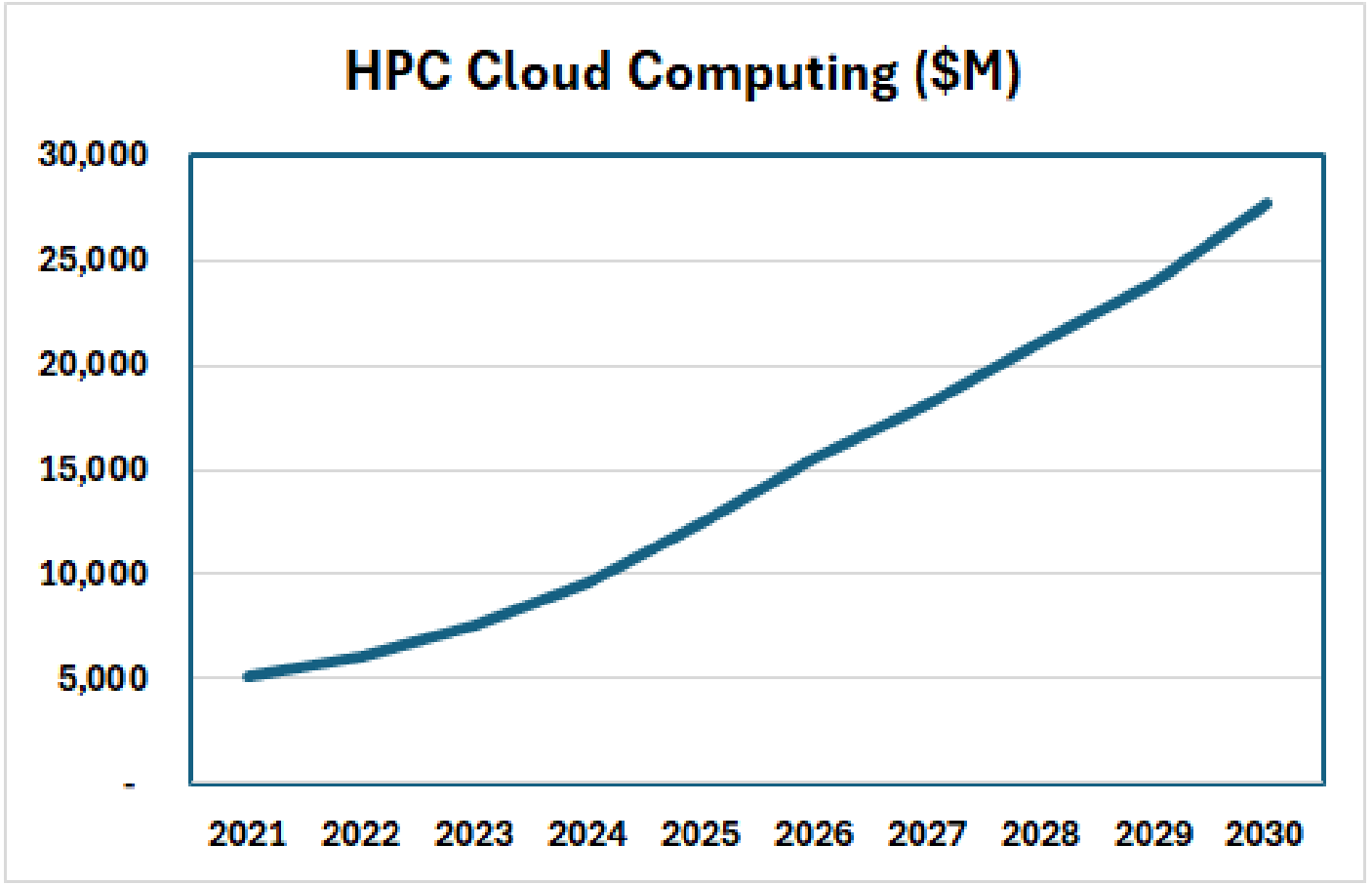
GPU Installations: Over 1 Million in 2025

In HPC & AI-for-Science = 33.3% CAGR over the last 5 years



Cloud Computing for HPC/AI-for-Science

Projected cloud HPC/AI spending to reach $30 billion by 2030
Grew 30% in 2025



HPC/AI-for-Science Market Should See Growth in 2026

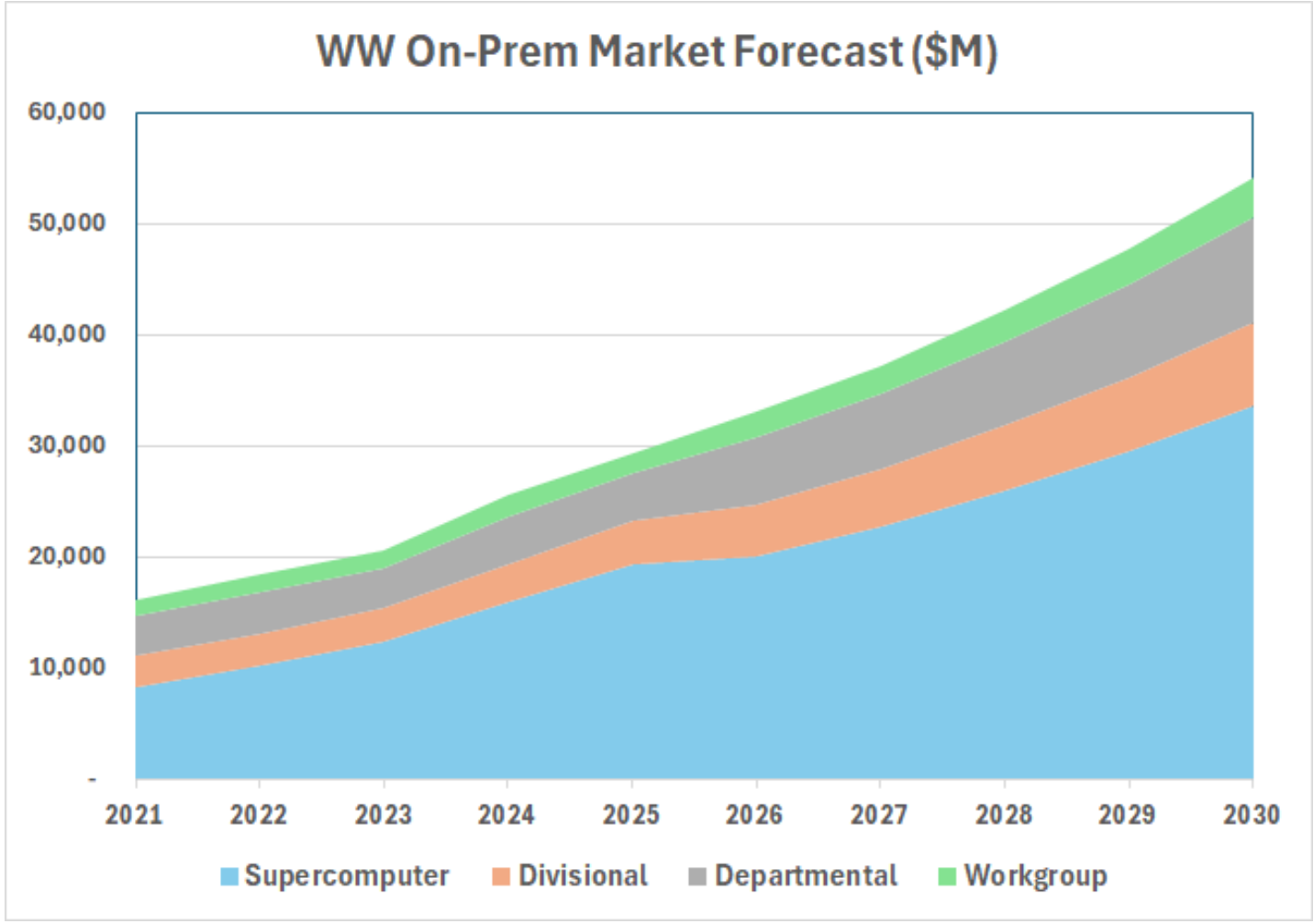
… but there are some major concerns

- **The global economic situation and changing trade rules could have a major impact to IT build outs in 2026**
- **Non-science AI systems are getting major attention and $$**
- **Supply chain issues are still impacting installations (e.g., GPUs, memory & SSDs)**
- **Exascale system acceptances are seeing delays**
- **The lower end of the on-premises market continues to struggle**

- **Growth drivers include:**
  - New use cases especially in AI are providing new areas for users to advance their research
  - Countries and companies around the world continue to recognize the value of being innovative and investing in R&D to advance society, grow revenues, reduce costs, and become more competitive

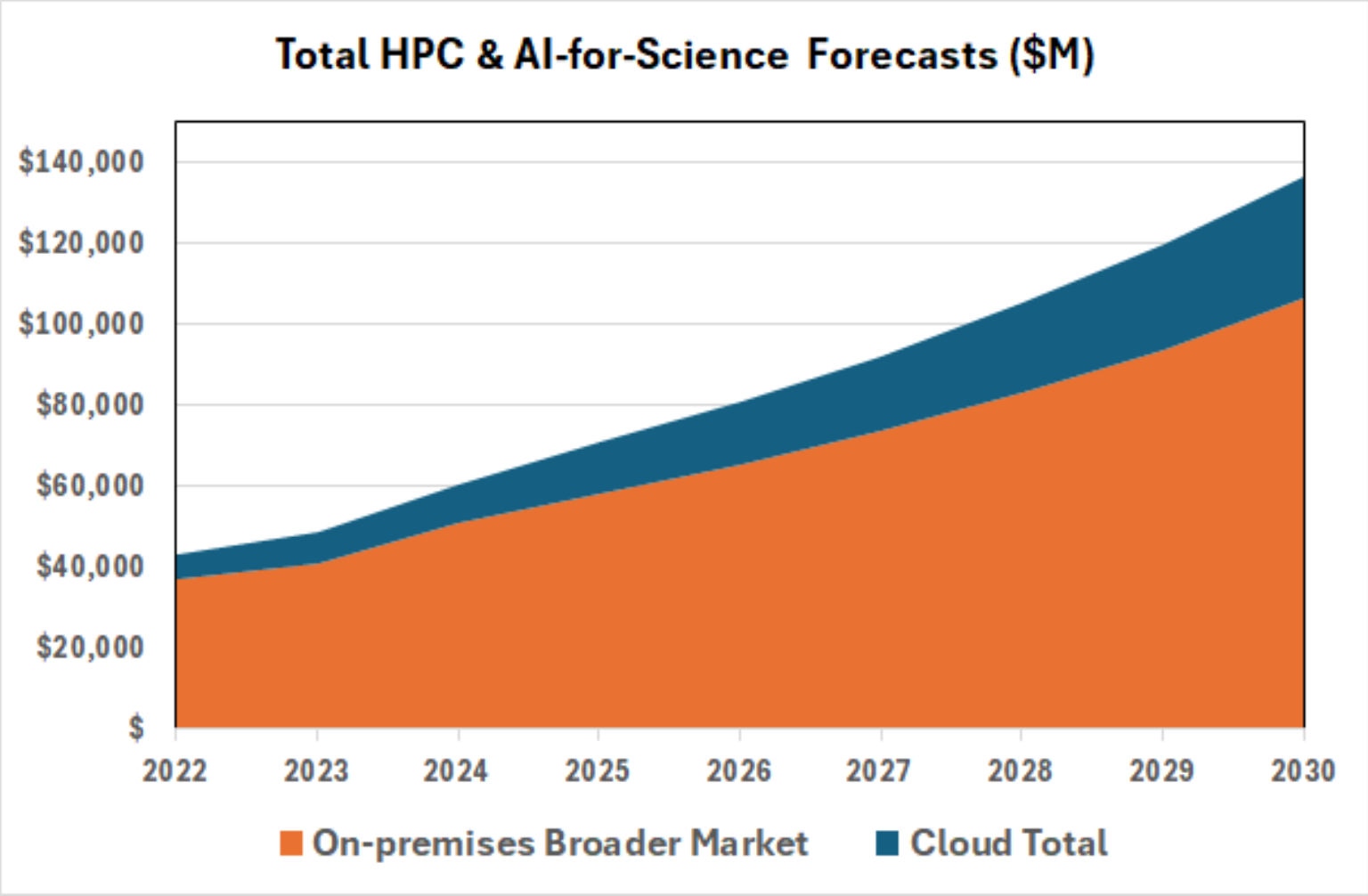
HPC/AI-for-Science On-Prem Server Forecast

On-prem servers are projected to reach $54 billion by 2030
Broader HPC/AI on-prem market is projected to exceed $100 billion



HPC & AI-for-Science On-Prem Plus Cloud Forecast

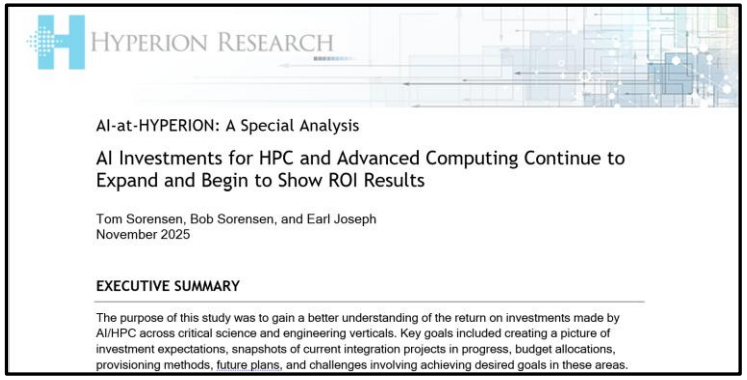
The overall market is projected to exceed $135 billion by 2030
Doubling over the next 5 years



Trends in AI-for-Science

Success in Using AI-for-Science

"AI Investments for HPC and Advanced Computing Continue to Expand and Begin to Show ROI Results"


HYPERION RESEARCH

AI-at-HYPERION: A Special Analysis

AI Investments for HPC and Advanced Computing Continue to Expand and Begin to Show ROI Results

Tom Sorensen, Bob Sorensen, and Earl Joseph
November 2025

EXECUTIVE SUMMARY

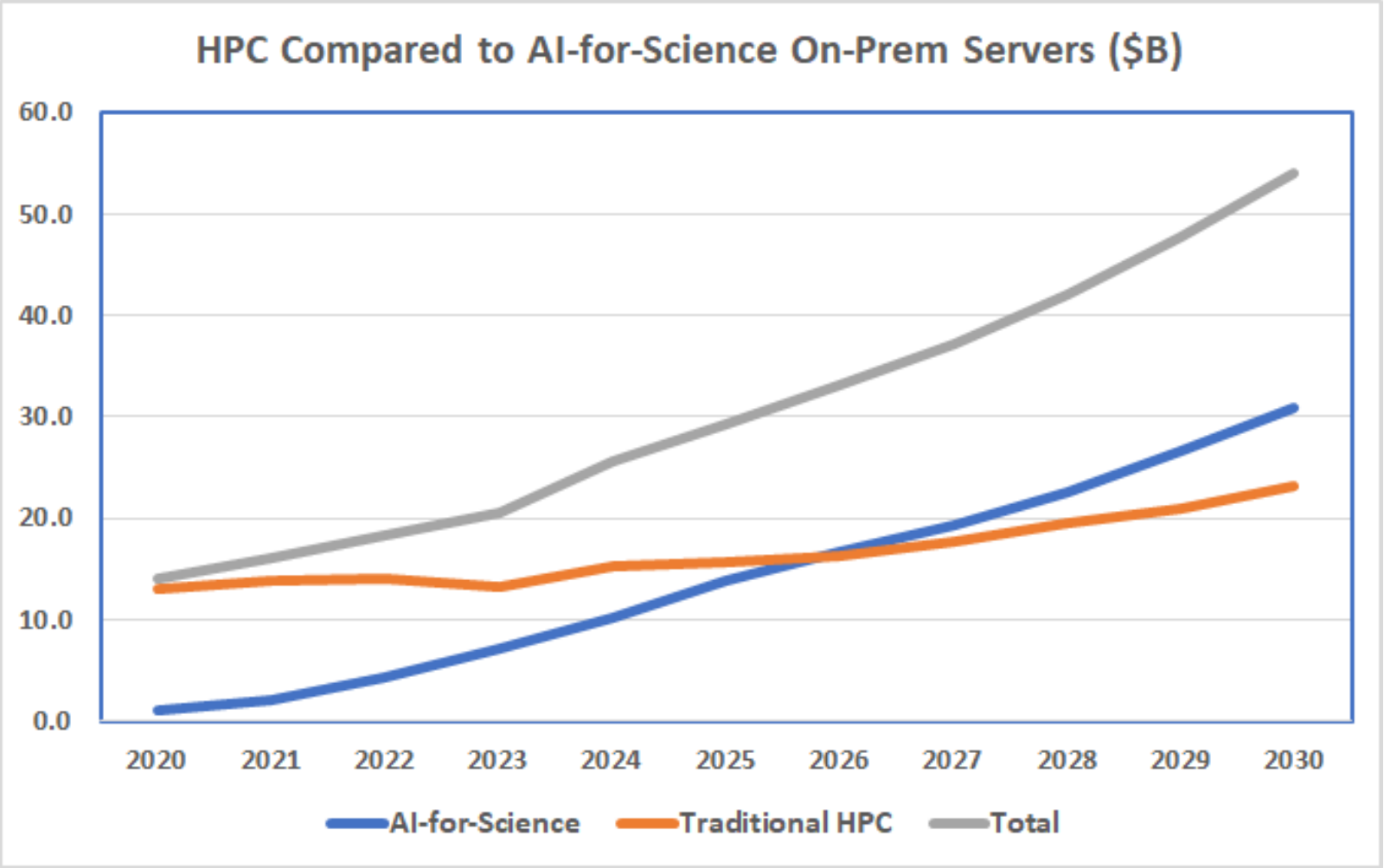
The purpose of this study was to gain a better understanding of the return on investments made by AI/HPC across critical science and engineering verticals. Key goals included creating a picture of investment expectations, snapshots of current integration projects in progress, budget allocations, provisioning methods, future plans, and challenges involving achieving desired goals in these areas.

Highlights from the study:

- **75.8% of the sites felt that their AI projects met or exceeded expectations**
- **74.9% of respondents indicated plans to moderately or significantly expand generative AI to support HPC workloads**
  - Less than 3% expect to contract their use of AI, none of which would characterize that contraction as significant
- **Roughly 40% of respondents are already using agentic AI models**
- **However, technical challenges continue to bring hesitation when it comes to broad adoption:**
  - Hallucinations, lack of explainability, and integration complexity are persistent concerns
  - This signals a transition from reactive adoption to more measured, application-specific onboarding

Traditional HPC & AI-for-Science

AI-for-Science servers are projected to reach $31 billion by 2030
Total HPC & AI-for-Science to reach $54 billion by 2030



AI-for-Science Servers by Region

North America is projected to grow at a high rate 18.5% CAGR
Followed by Europe at 17.4% CAGR

Hyperion Research AI-for-Science Server Forecast: By Region ($ Billions)								
	2024	2025	2026	2027	2028	2029	2030	CAGR 2025-30
North America	4.76	6.51	8.01	9.46	11.06	13.12	15.22	18.5%
EMEA	2.84	3.79	4.58	5.27	6.17	7.31	8.48	17.4%
Asia/Pacific w/o Japan	1.92	2.54	3.02	3.43	3.96	4.63	5.30	15.9%
Japan	0.56	0.74	0.88	1.00	1.14	1.32	1.49	15.2%
Rest-of-World	0.15	0.20	0.23	0.26	0.30	0.34	0.38	14.0%
Total Server Revenue	10.2	13.8	16.7	19.4	22.6	26.7	30.9	17.5%
Source: Hyperion Research, May 2026								

Note: these numbers do not yet include Genesis and European AI factories.

AI-for-Science Servers by Verticals

Government Lab is the leading sector: over $12 billion in 2030

Hyperion Research AI-for-Science Server Forecast: By Verticals ($ Billions)								
	2024	2025	2026	2027	2028	2029	2030	CAGR 2025-30
Bio-Sciences	1.30	1.76	2.10	2.36	2.67	3.06	3.43	14.2%
CAE	0.77	1.04	1.24	1.39	1.57	1.80	2.02	14.2%
Chemical Engineering	0.10	0.14	0.16	0.18	0.21	0.24	0.27	14.2%
DCC & Distribution	0.46	0.62	0.74	0.83	0.94	1.08	1.21	14.2%
Economics/Financial	0.36	0.52	0.62	0.69	0.79	0.90	1.01	14.2%
EDA	0.15	0.21	0.25	0.28	0.32	0.37	0.41	14.2%
Geosciences	0.18	0.24	0.28	0.32	0.36	0.41	0.46	14.2%
Defense	1.34	1.86	2.28	2.67	3.15	3.75	4.38	18.7%
Government Lab	3.66	4.79	5.96	7.19	8.69	10.60	12.63	21.4%
University/Academic	1.58	2.15	2.56	2.88	3.25	3.73	4.18	14.2%
Weather	0.23	0.31	0.36	0.41	0.46	0.53	0.60	14.2%
Other	0.10	0.14	0.17	0.19	0.21	0.25	0.28	14.2%
Total Server Revenue	**10.2**	**13.8**	**16.7**	**19.4**	**22.6**	**26.7**	**30.9**	**17.5%**

Source: Hyperion Research, May 2026

2026 MCS Study Findings: AI Strategies

94.8% are using or will soon be using AI/LLMs
But 8.2% plan to stop using them

Characterizing AI/LLM Strategy	
Q. Which of the following best describes your generative AI/LLM strategy?	
Strategy Characterization	**Percentage of Sites**
We use generative AI/LLMs today and plan to continue using it over the next 12-18 months	71.1%
We do not use generative AI/LLMs today but plan to start in the next 12-18 months	15.5%
We use generative AI/LLMs today but plan to stop using it within the next 12-18 months	8.2%
We do not use generative AI/LLMs today and do not plan to within the next 12-18 months	5.2%
n = 97	
Source: Hyperion Research, 2026	

Study Findings: Barriers to AI Growth

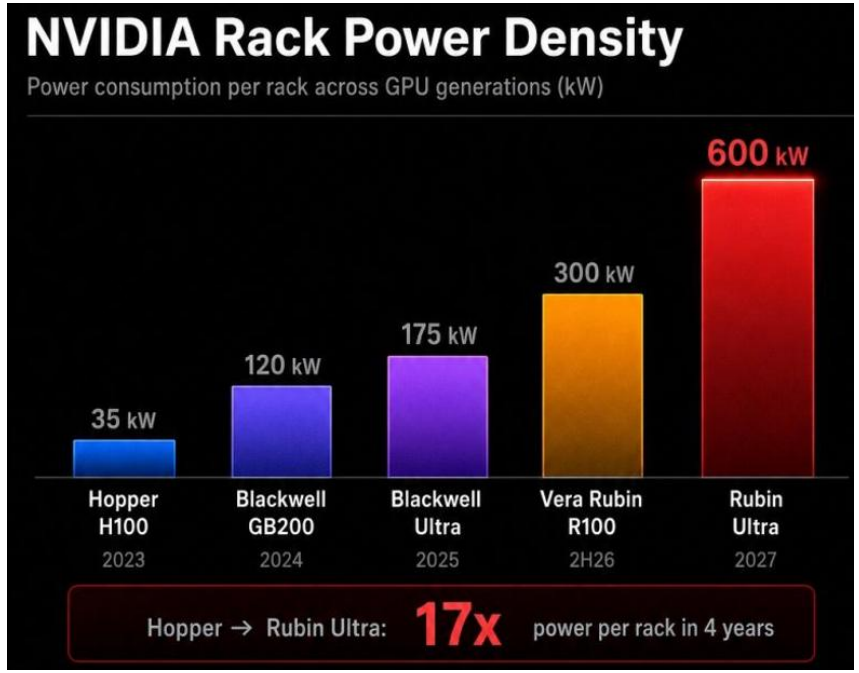
Expertise, quality of training data, dealing with complexity and high costs are the top barriers

Barriers to AI Capability Growth	
Q. Are any of the following barriers to furthering your AI capabilities? Please select all that apply:	
Barrier	**Percentage of Sites**
Level of in-house AI expertise	36.7%
Quality of available training data	35.0%
Complexity with integrating AI models into existing HPC workloads	30.0%
High/uncertain development costs	28.3%
Concerns with technical issues (i.e., scaling, hallucinations, ease of maintenance)	21.7%
High/uncertain operational costs	20.8%
Scale of available training data	19.2%
Access to specialized hardware	17.5%
The technology is moving too fast for credible assessment of value	17.5%
Lack of demonstrated return on investment	16.7%
Level of AI vendor or 3rd party expertise	10.8%
Access to specialized software	10.0%
Confusion/uncertainty with vendor selection	8.3%
Other (please specify)	3.3%
n = 120	
Notes: Respondents could select multiple options	
Source: Hyperion Research, 2026	

Market Changes:
Nodes Per System

Rack Power Requirements

Drives major increases in the price per rack



WHAT 600kW PER RACK MEANS FOR INFRASTRUCTURE

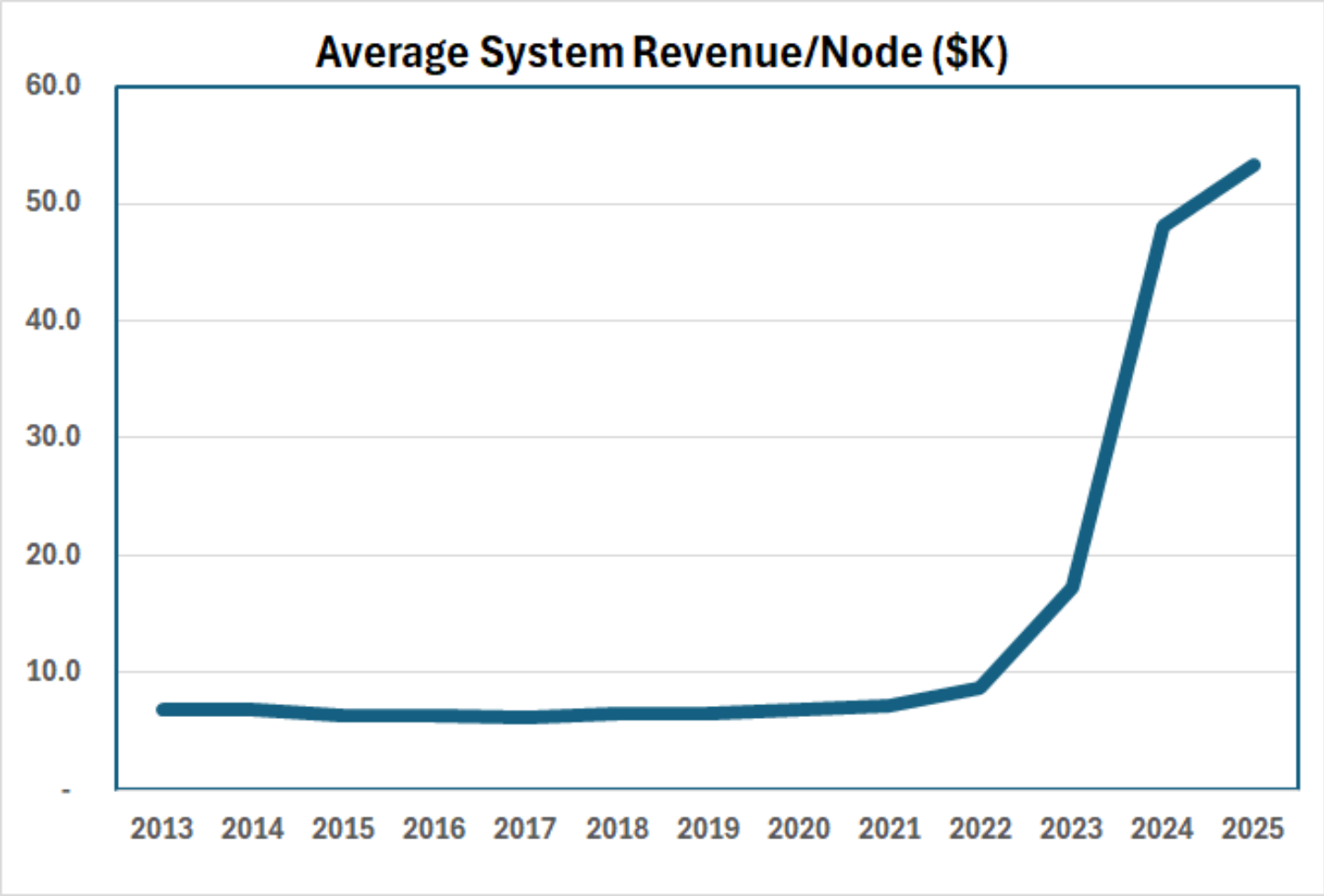
	CONSTRAINT	HOPPER ERA (35kW)	RUBIN ULTRA (600kW)
❄	Cooling	Air Cooled	Direct-to-Chip Liquid Only
⚡	Power per 10K racks	350 MW	6,000 MW (6 GW)
	Transformer demand	Baseline	~17x baseline

The bottleneck isn't chips. It's everything downstream: cooling, transformers, grid connections, and time. Jensen said NVIDIA will be "supply constrained through the entire life of Vera Rubin." The constraint he's talking about isn't silicon. It's infrastructure.

Prejean Consulting | May 2026

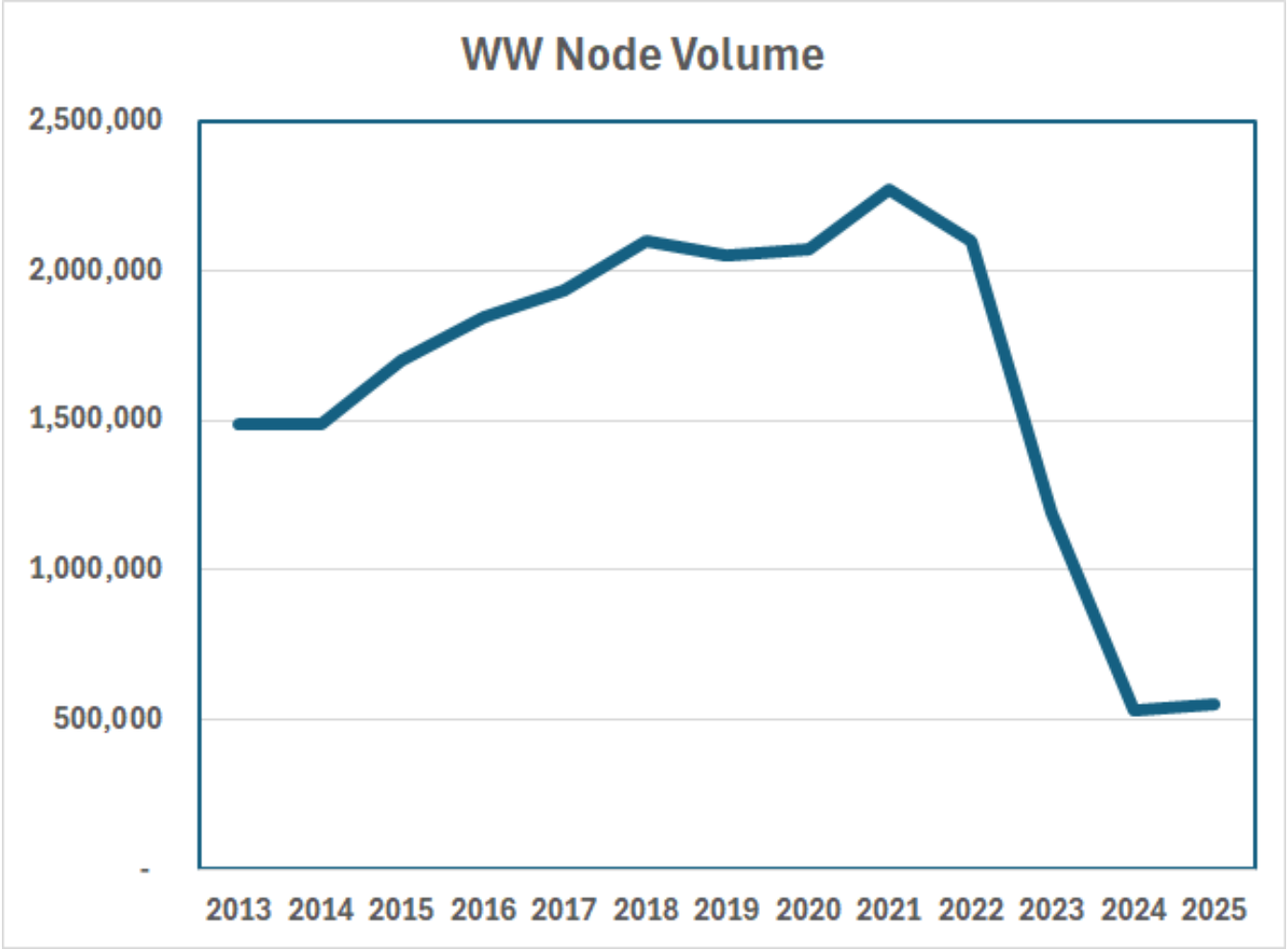
A Major Market Shift: High Node Costs

Driven by more systems having GPUs, higher GPU costs and higher node costs (memory, liquid cooling, higher power)



A Major Market Shift: Fewer Node Purchases

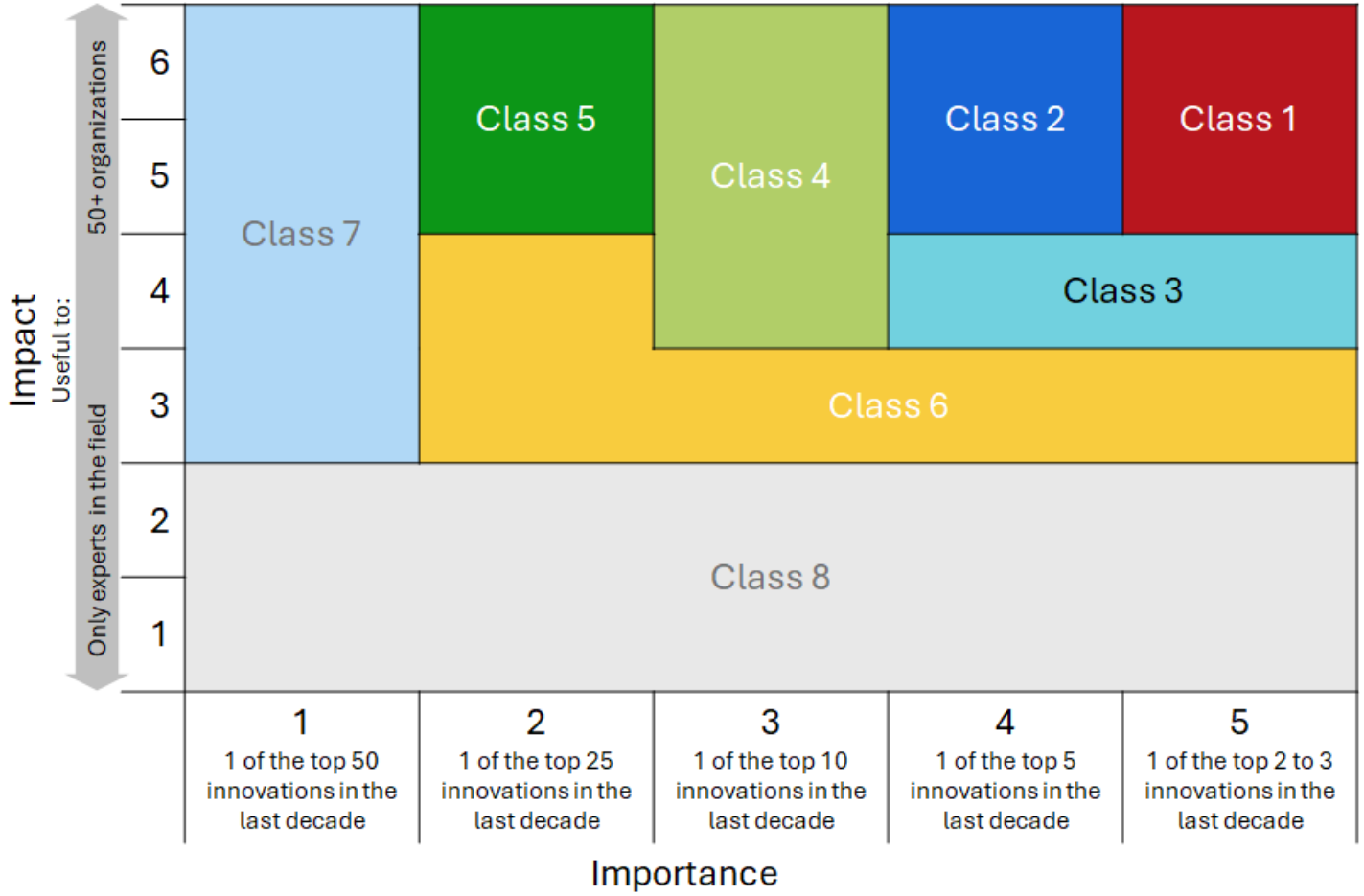
Driven by more systems having GPUs, higher GPU costs and higher node costs (memory, liquid cooling, higher power)



Measuring and Comparing Leadership Computing

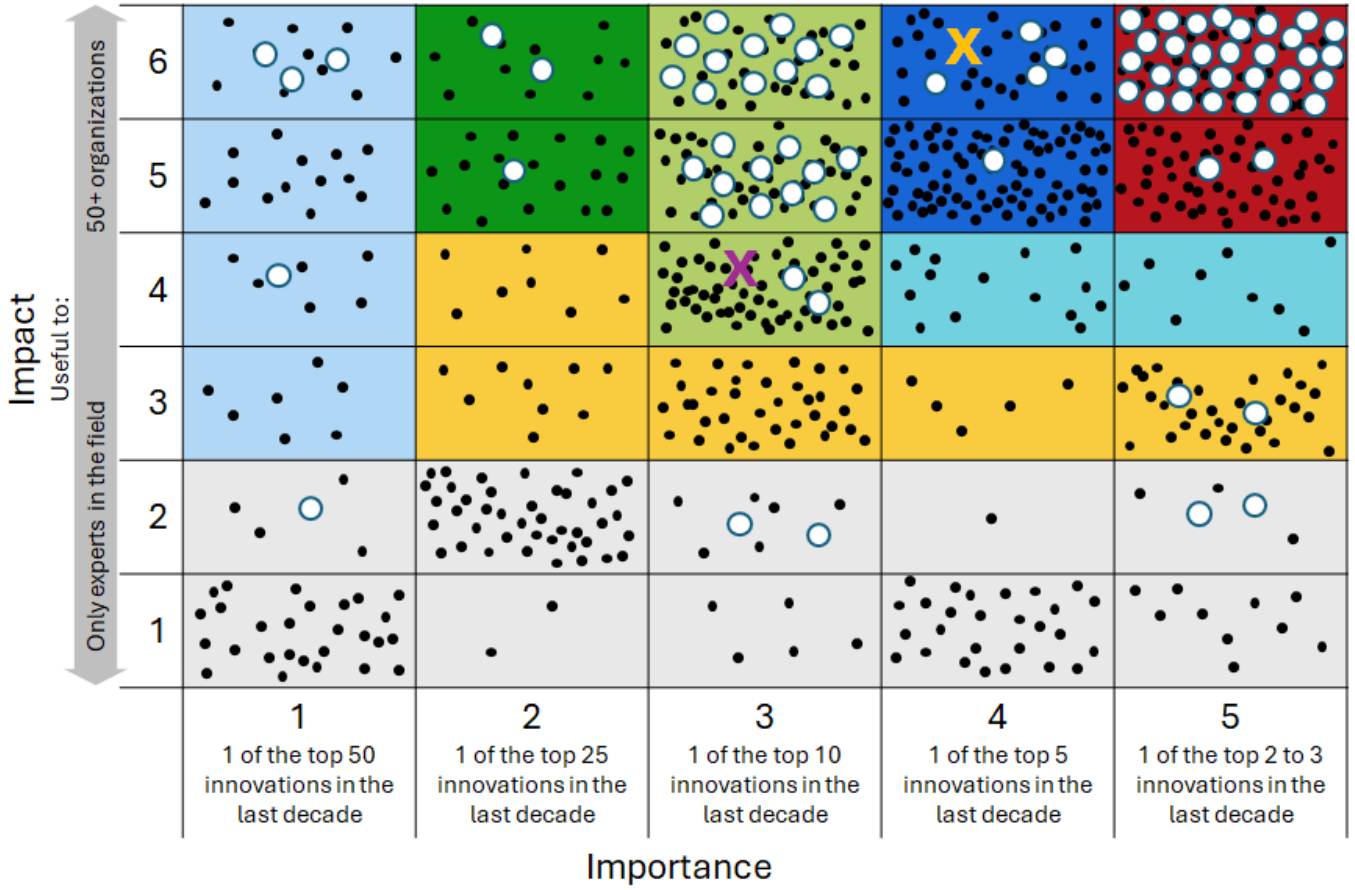
One Way to Show the Value of Leadership Computing

Using two scales: innovation importance level, and how broadly impactful are the results



Showing the Value of Leadership Computing: RIKEN

An example from a 2024 study compared to 650 other projects
When applied to key societal goals, the increased value generated is impressive



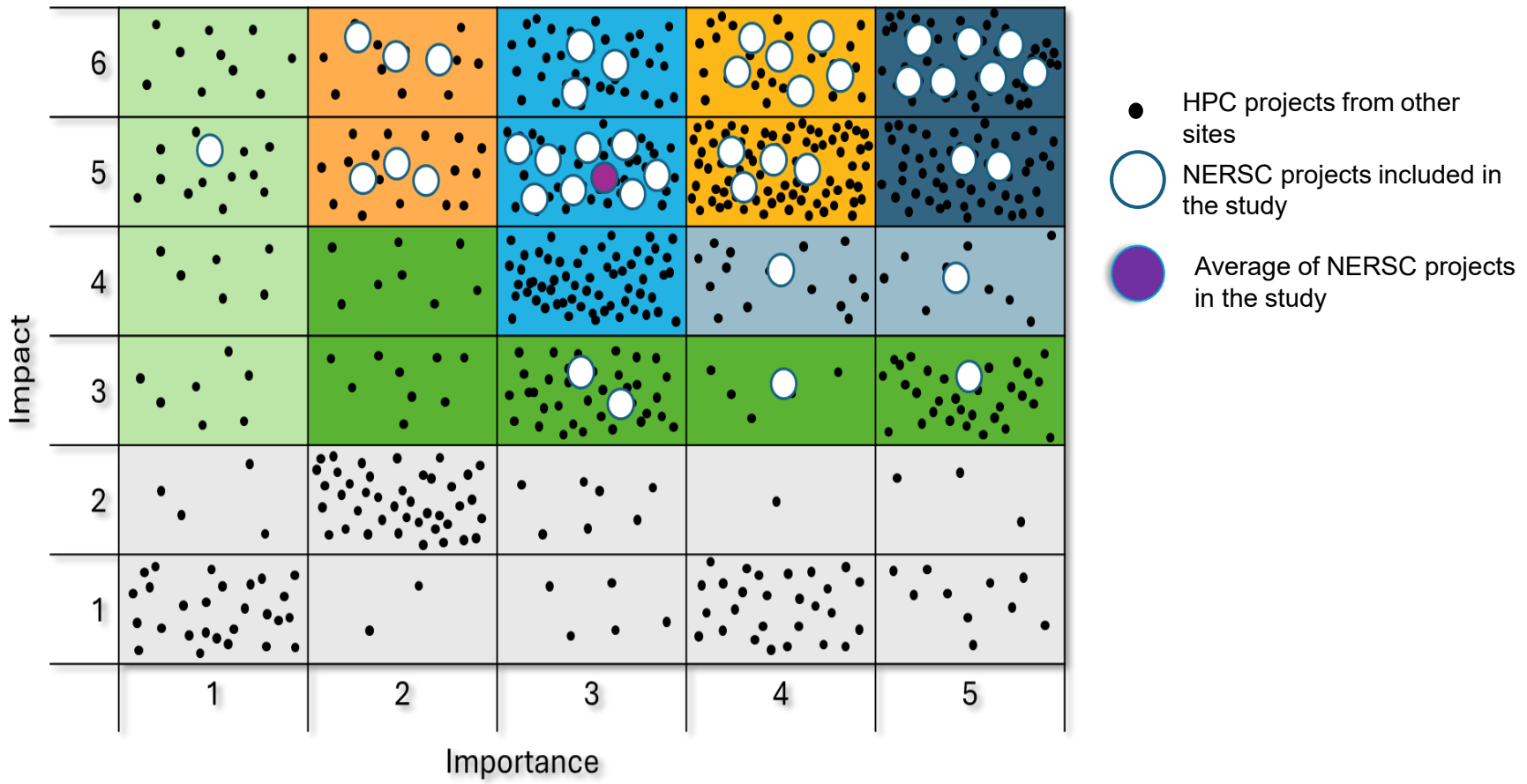
The K computer exceeded 9 times its investment.

If Fugaku delivers similar returns as the K computer, plus the added returns from addressing major societal issues, the returns could exceed 68 to 90 times the investment in the Fugaku system.

Showing the Value of Leadership Computing: NERSC

An example from a 2024 study compared to 650 other projects

Innovation Class Mapping: Showing Participating NERSC projects



Conclusions

Expecting strong growth, but there are some growing concerns…

- **2025 was a strong growth year**
  - AI-for-Science, GPUs, and cloud are high growth areas
  - New QC systems are being installed around the world
  - 2026 looks like a healthy growth year
- **New technologies are showing up large numbers:**
  - Agentic AI, Generative AI, smarter AI, LLMs and SLLs are fueling a new level of growth
  - Processors, AI hardware & software, memories, new storage approaches, etc.
  - The cloud has become a viable option for many HPC workloads
- **Storage will likely see major growth driven by AI, big data and the need for much larger data sets**
- **There are growing concerns around system costs, supply chains, power, AI impacts, data centers push-back, talent and political changes**

HYPERION RESEARCH

We Welcome Questions, Comments and Suggestions

Please contact us at:

info@hyperionres.com