



HYPERION RESEARCH

# Hyperion Research Market Update

October 2025

[www.HyperionResearch.com](http://www.HyperionResearch.com)  
[www.hpcuserforum.com](http://www.hpcuserforum.com)

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

# About Hyperion Research



([www.HyperionResearch.com](http://www.HyperionResearch.com) & [www.HPCUserForum.com](http://www.HPCUserForum.com))

## Hyperion Research Mission:

- Hyperion Research helps organizations make effective decisions and seize growth opportunities
  - By providing research and recommendations in high performance computing and emerging technology areas

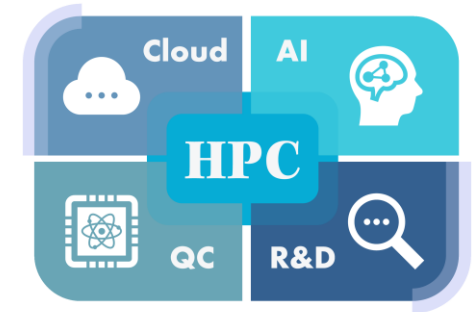
## HPC User Forum Mission:

- To improve the health of the HPC/AI/QC industry
  - Through open discussions, information sharing and initiatives involving HPC users in industry, government and academia along with HPC vendors and other interested parties

# Example Research Areas

([www.HyperionResearch.com](http://www.HyperionResearch.com) & [www.HPCUserForum.com](http://www.HPCUserForum.com))

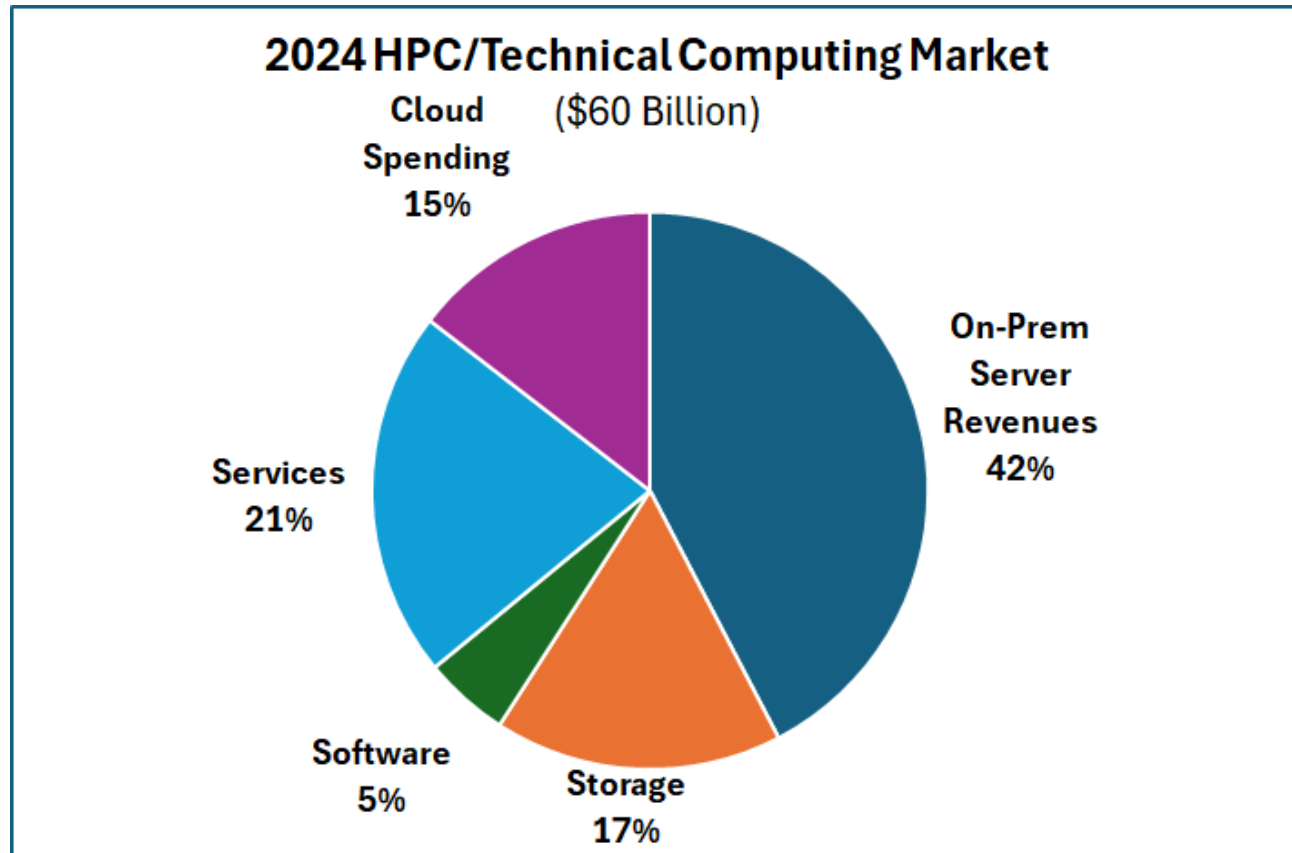
- **Traditional HPC**
- **AI, ML, DL, LLMs, Graph**
- **Cloud Computing**
- **Storage & Data**
- **Interconnects**
- **Software & Applications**
- **ROI and Scientific Returns from HPC**
- **Power & Cooling**
- **Tracking all Processor Types & Growth rates**
- **Quantum Computing**
- **R&D and Engineering -- all types**
- **Edge Computing**
- **Supply Chain Issues**
- **Sustainability**



# HPC/AI Market Update

# 2024 Was a Strong Growth Year

*The highest growth in over two decades (23.4%)!*



- **23.4% growth in on-premises servers**
- **21.3% growth in the use of clouds**
- **Over \$60 billion in total spending**

# 2024 HPC/AI Market By Segment

*The highest growth in over two decades (23.4%)!*

2024 HPC/AI Market By Segment		
2024 New Segments	2024 Server Revenues	2024 Market Shares
Leadership Computers (>\$150M)	1,190	4.7%
Supercomputers (\$10M-\$150M)	6,921	27.3%
Large HPC (\$1M-\$10M)	7,078	27.9%
Medium HPC (\$250K-\$1M)	3,985	15.7%
Entry HPC (<\$250K)	6,159	24.3%
<b>Total</b>	<b>25,332</b>	<b>100.0%</b>
<i>Source: Hyperion Research, 2025</i>		

# 2024 HPC/AI Market By Region

*The highest growth in over two decades (23.4%)!*

2024 HPC/AI Market By Region		
2024 New Segments	2024 Server Revenues	2024 Market Shares
North America	13,421	53.0%
EMEA	6,168	24.3%
Asia/Pacific (All)	5,467	21.6%
Rest of World	276	1.1%
<b>Total</b>	<b>25,332</b>	<b>100.0%</b>
<i>Source: Hyperion Research, 2025</i>		

# 2024 HPC/AI Market By Vertical

*The highest growth in over two decades (23.3%)!*

<b>WW High-Performance Systems Revenue by Applications</b>			
	<b>2023</b>	<b>2024</b>	<b>2023 to 2024 Growth</b>
<b>Bio-Sciences</b>	\$1,883	\$2,279	21.0%
<b>CAE</b>	\$2,319	\$2,729	17.7%
<b>Chemical Engineering</b>	\$236	\$301	27.5%
<b>DCC &amp; Distribution</b>	\$1,143	\$1,389	21.5%
<b>Economics/Financial</b>	\$1,044	\$1,323	26.7%
<b>EDA / IT / ISV</b>	\$1,196	\$1,480	23.7%
<b>Geosciences</b>	\$1,300	\$1,543	18.6%
<b>Mechanical Design</b>	\$058	\$061	4.4%
<b>Defense</b>	\$2,151	\$2,563	19.2%
<b>Government Lab</b>	\$4,446	\$6,114	37.5%
<b>University/Academic</b>	\$3,482	\$4,012	15.2%
<b>Weather</b>	\$940	\$1,127	20.0%
<b>Other</b>	\$350	\$412	17.6%
<b>Total Server Revenue</b>	<b>\$20,550</b>	<b>\$25,333</b>	<b>23.3%</b>

*Source: Hyperion Research, 2025*

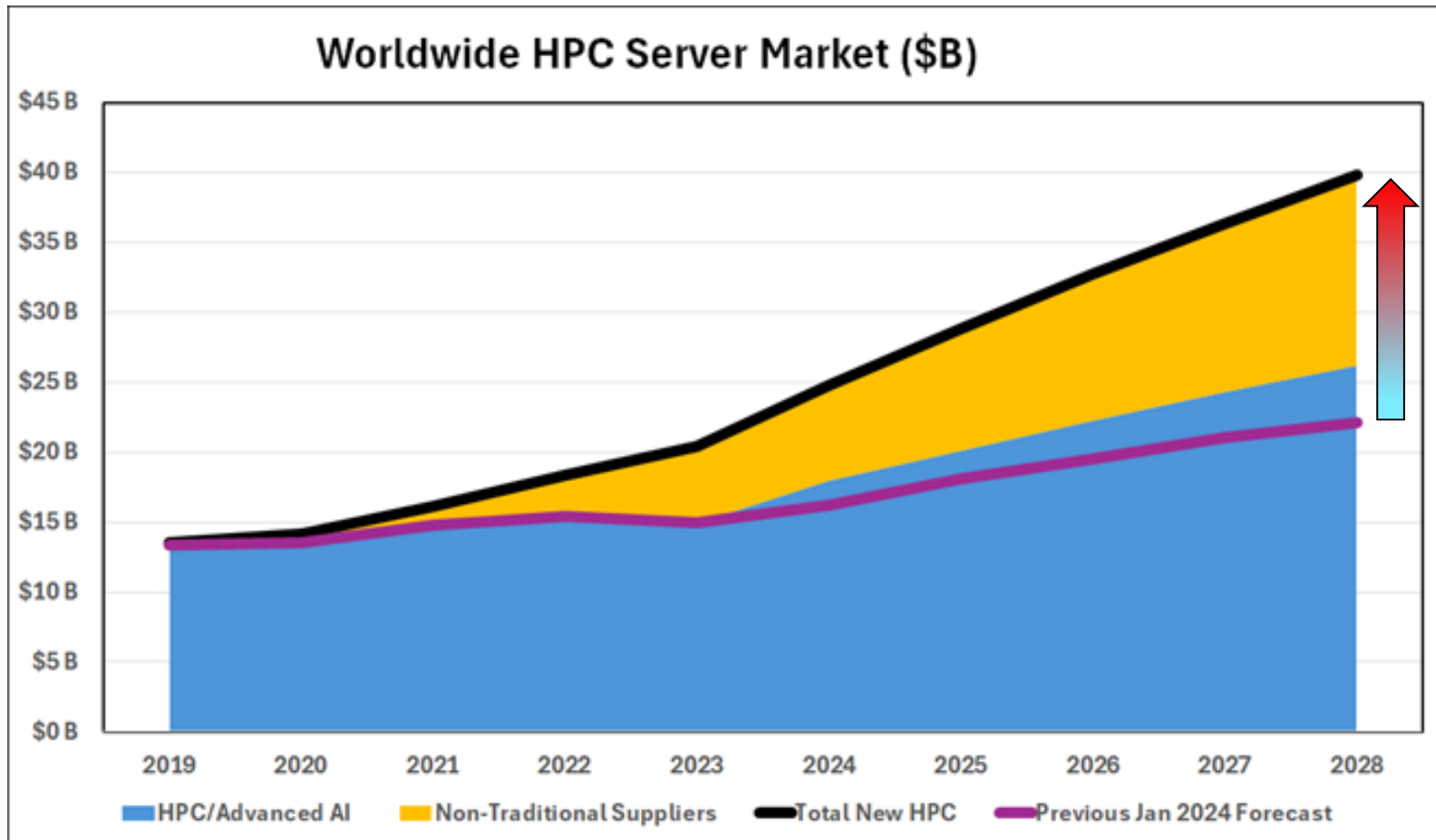
# The HPC/AI Market Should See Growth in 2025

*... but there are some major concerns*

- **The global economic situation and changing trade rules could have a major impact to IT build outs in 2025**
- **Supply chain issues are still impacting installations (e.g., GPUs)**
- **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/LLMs/Generative AI/Smarter 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

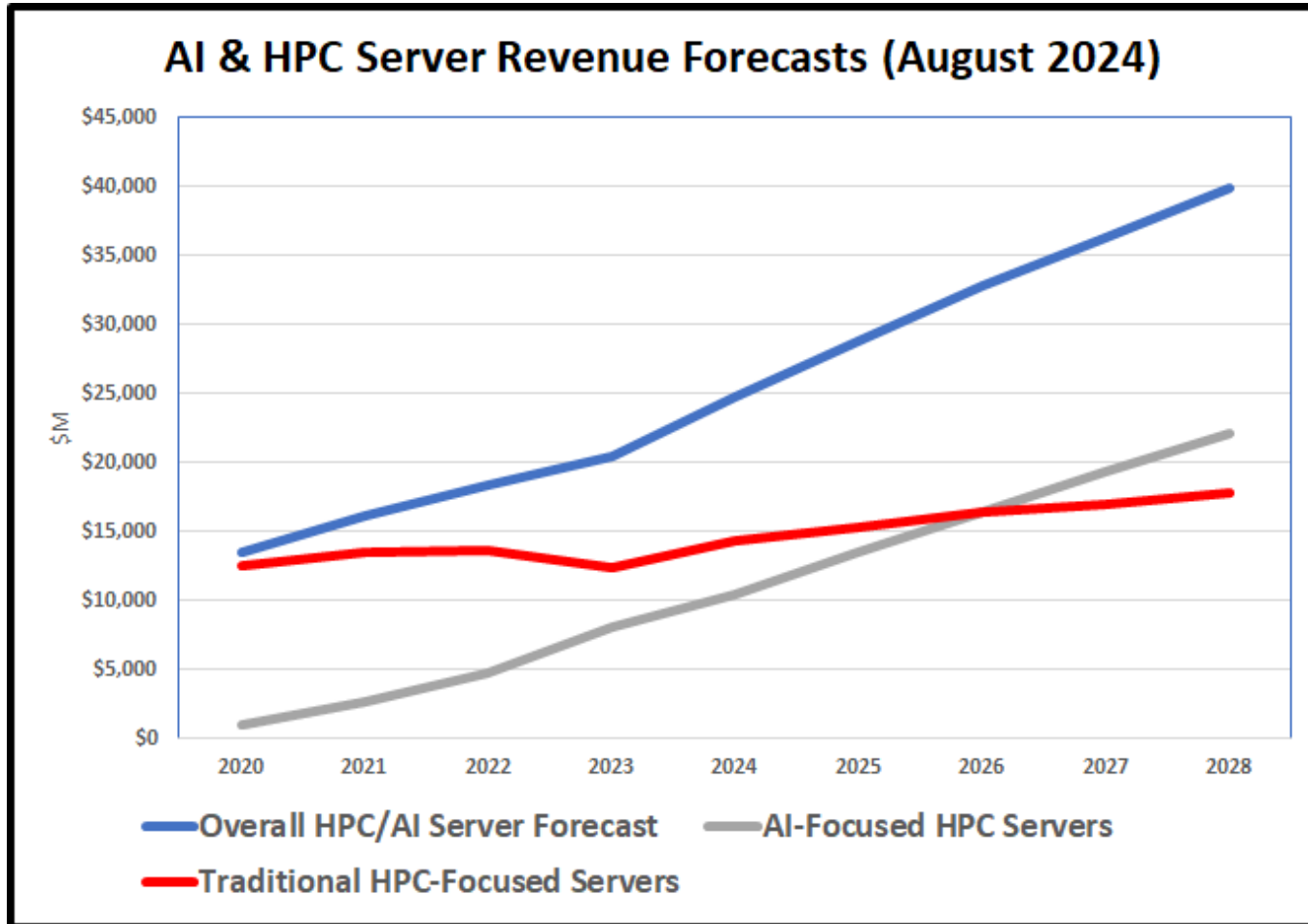
# AI Has Ignited Growth

- *Hyperion Research announced a 36.7% increase in the HPC/AI server market size (now growing at 15% CAGR)*
- *Added tracking of non-traditional AI/HPC suppliers*



# HPC Compared to AI-centric Servers

*Many servers are running both traditional HPC and AI Workloads*

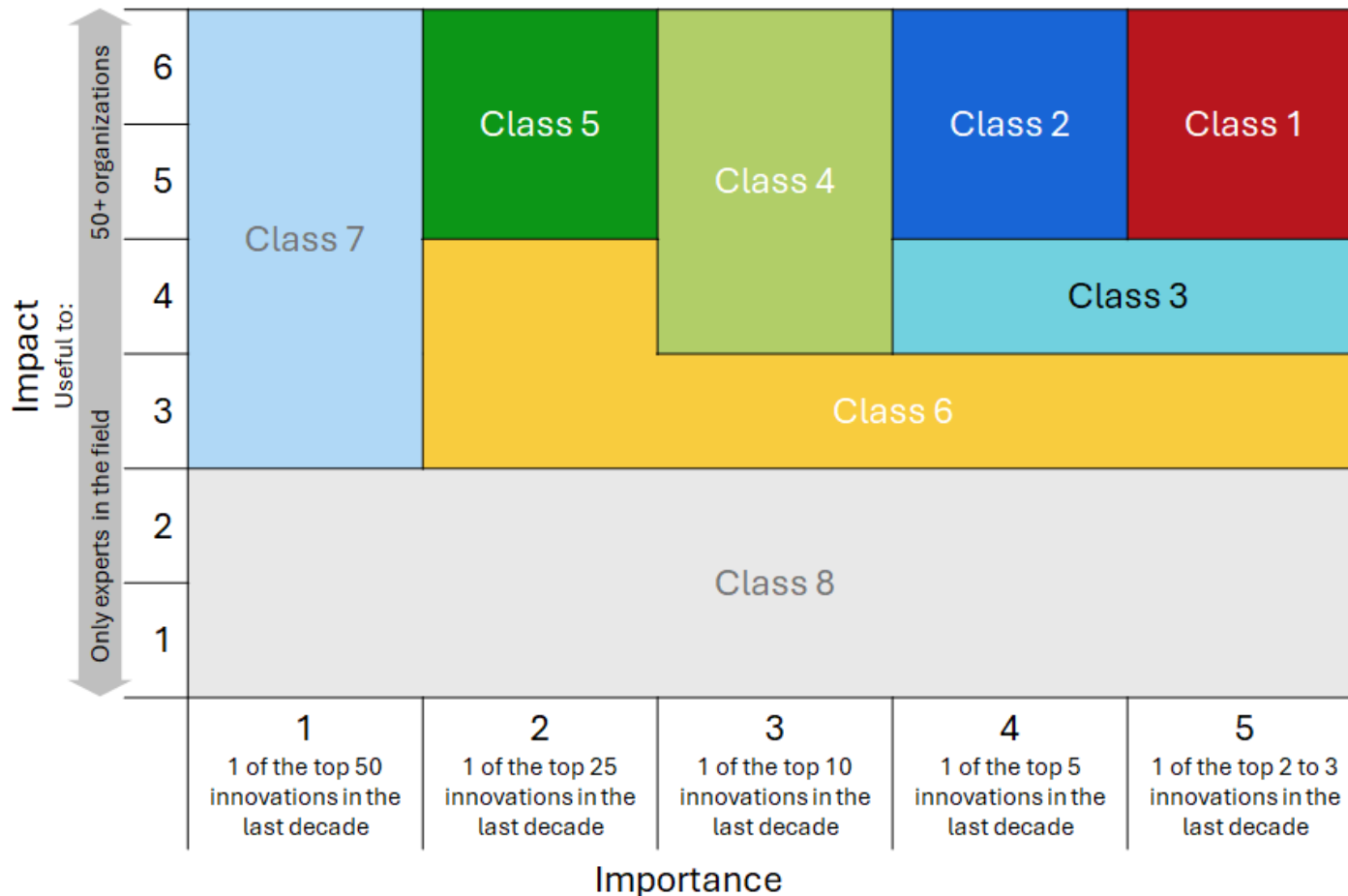


Note: AI systems may still run some traditional HPC jobs (<50% of workload).  
Likewise, traditional HPC systems often run some AI jobs (<50% of workload).

# Measuring and Comparing Leadership Computing

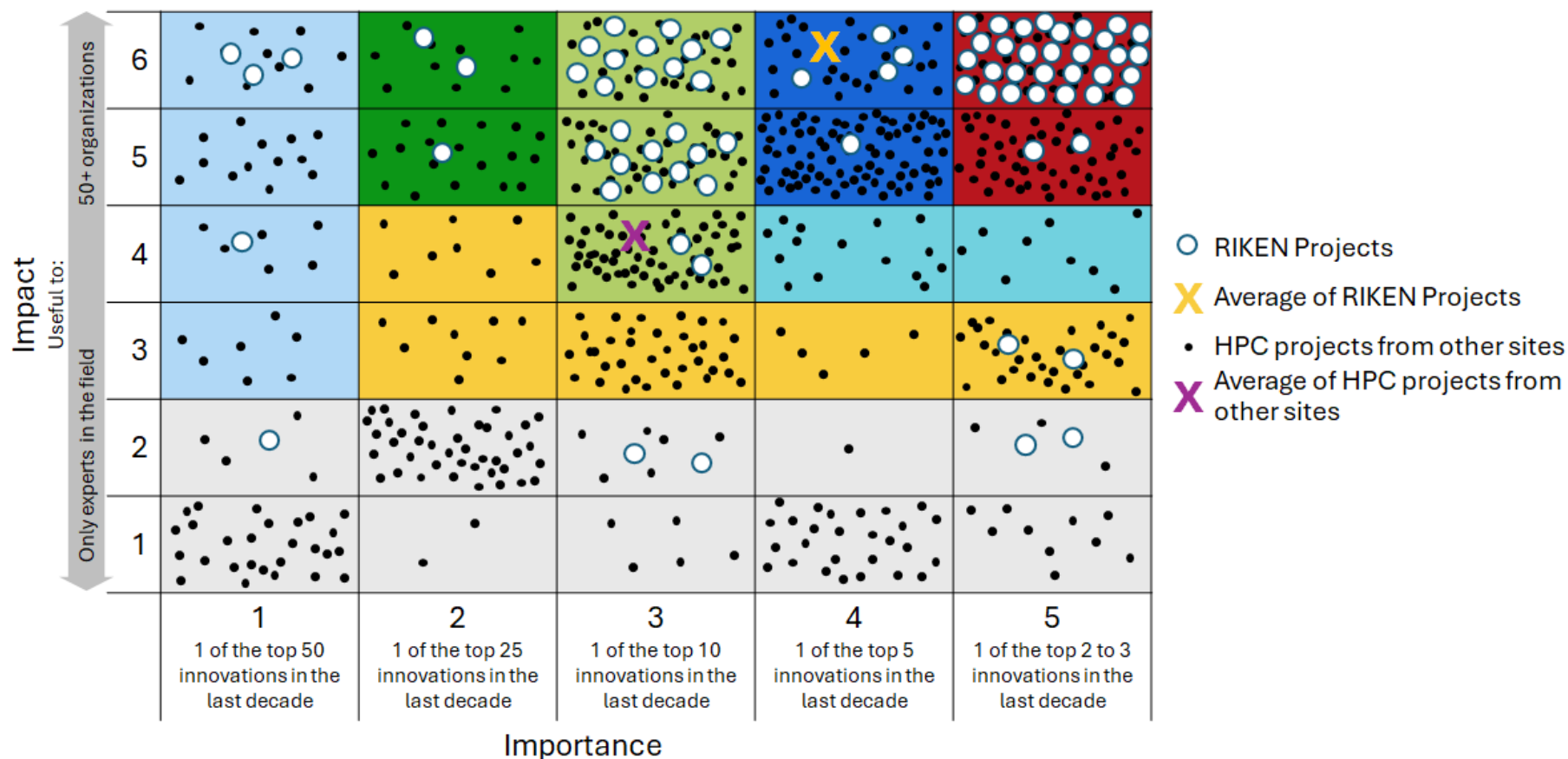
# A New Way to Show the Value of Leadership Computing

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



# A New Way to Show the Value of Leadership Computing - RIKEN

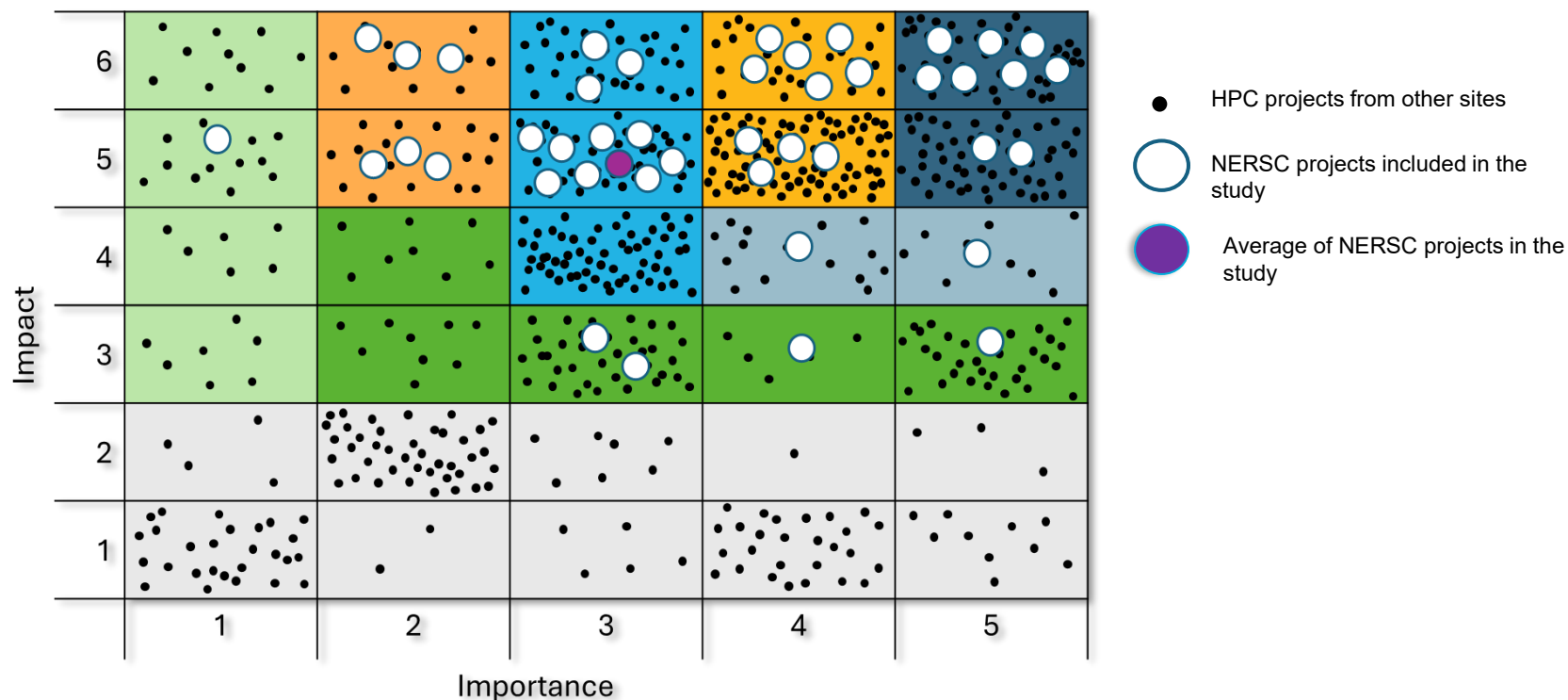
*An example from a 2024 study compared to 650 other projects*



# A New Way to Show the Value of Leadership Computing - NERSC

*An example from a 2024 study compared to 650 other projects*

Innovation Class Mapping: Showing Participating NERSC projects



# Some Results From Our AI Studies

# AI Growth

*Most sites are already running AI workloads, and plan to increase the scale this year*

52) Does your organization currently run AI/ML/DL/LLM workloads, or plan to in the near future?

Response	All Sectors		Academia		Government		Industry	
	# Responses	% of total	# Responses	% of total	# Responses	% of total	# Responses	% of total
Yes	103	96%	13	93%	16	100%	74	96%
No	4	4%	1	7%			3	4%
Respondents	107		14		16		77	
Total surveyed	107		14		16		77	

53) Have your AI plans changed in the past 12 months? If so, how? Select all that apply

Response	All Sectors		Academia		Government		Industry	
	# Responses	% of respondents	# Responses	% of respondents	# Responses	% of respondents	# Responses	% of respondents
We have increased the scale of our AI workloads in the past 12 months	77	75%	12	92%	11	69%	54	73%
We have moved up the timeline of AI adoption in the past 12 months	30	29%	2	15%	5	31%	23	31%
We are now employing different methodologies than previously planned in	14	14%			2	13%	12	16%
Our AI plans have not changed in the past 12	7	7%			1	6%	6	8%
Total responses selected	128		14		19		95	
Respondents	103		13		16		74	
Total surveyed	107		14		16		77	

# Using Clouds for AI

*Clouds are a viable option for many AI workloads*

55) Please approximate the portion of your AI/ML/DL/LLM workload done in the cloud compared to on-premises (based on CPU hours):

Response	All Sectors	
	# Responses	% of total
0% on-premises and 100% in the cloud	4	4%
10% on-premises and 90% in the cloud	10	9%
25% on-premises and 75% in the cloud	14	13%
50% on-premises and 50% in the cloud	14	13%
75% on-premises and 25% in the cloud	22	21%
90% on-premises and 10% in the cloud	17	16%
100% on-premises and 0% in the cloud	22	21%
Respondents	103	
Total surveyed	107	

# Inferencing

*Inferencing is over 50% of the AI workload at 85% of the sites in the study – BUT training is growing faster*

56) Please approximate the portion of your AI/ML/DL/LLM workload that is training compared to inferencing (based on

	Response	All Sectors	
		# Responses	% of total
	0% training and 100% inferencing	1	1%
	10% training and 90% inferencing	4	4%
	25% training and 75% inferencing	11	10%
	50% training and 50% inferencing	24	22%
	75% training and 25% inferencing	38	36%
	90% training and 10% inferencing	20	19%
	100% training and 0% inferencing	5	5%
	Respondents	103	
	Total surveyed	107	

57) Which do you expect to grow faster over the next 12 to 18 months (in CPU hours)?

	Response	All Sectors	
		# Responses	% of total
	Training	67	63%
	Inferencing	36	34%
	Respondents	103	
	Total surveyed	107	

# Generative AI

## *Generative AI is growing quickly*

60) Which of the following best describes your generative AI/LLM strategy?			
	Response	All Sectors	
		# Responses	% of total
	We use generative AI/LLMs today and plan to continue using it over the next 12-18 months	77	72%
	We do not use generative AI/LLMs today but plan to start in the next 12-18 months	14	13%
	We use generative AI/LLMs today but plan to stop using it within the next 12-18 months	9	8%
	We do not use generative AI/LLMs today and do not plan to within the next 12-18 months	3	3%
	<b>Respondents</b>	<b>103</b>	
	<b>Total surveyed</b>	<b>107</b>	

# AI Frameworks

*Sites are using a broad set of frameworks to develop their AI applications*

63) What AI/ML/DL/LLM/data intensive frameworks do you use for your HPC workloads? Select all that apply

Response	All Sectors	
	# Responses	% of respondents
PyTorch	64	62%
Tensorflow	52	50%
Cuda	44	43%
Jupyter Notebooks	40	39%
OpenAI GPT	35	34%
SciKit Learn	28	27%
Spark	28	27%
SQL	23	22%
Hadoop	20	19%
Meta Llama	17	17%
Homegrown/In-House	16	16%
Keras	16	16%
Microsoft Cognitive Toolkit	12	12%
Torch	12	12%
Anthropic Claude	10	10%
Hive	9	9%
Caffe/Caffe 2	7	7%
IBM IIS/Watson/WebSphere	7	7%
Flink	6	6%
HPCC	6	6%
Theano	6	6%
Oracle WebLogic	4	4%
Other	9	9%

# Important Factors for Success

*More experts are the most needed for AI research*

66) Please select the top 2 in importance to your AI research:

Response	All Sectors	
	# Responses	% of respondents
AI expertise	41	40%
Domain expertise	36	35%
Breakthroughs in science and engineering	23	22%
Incorporating AI into traditional mod/sim workloads	23	22%
Cleaning up data	16	16%
Increasing the scale of computing	15	15%
Conducting new types of research	14	14%
Computer science expertise	13	13%
Expanding applications	9	9%
Benchmarking the capabilities of AI systems	8	8%
New methodologies	6	6%
Automating tasks	1	1%
Business case and benefits case for investing in AI	1	1%
Total # of responses	206	

# Barriers to Using AI

*Data quality, experts and the scale of data are the largest barriers today*

67) Are any of the following barriers to furthering your AI capabilities?

Response	All Sectors	
	# Responses	% of respondents
Quality of available training data	55	53%
Level of in-house AI expertise	40	39%
Scale of available training data	40	39%
Complexity with integrating AI models into existing HPC workloads	32	31%
Diversity of available training data	31	30%
Access to specialized hardware	29	28%
High/uncertain development costs	23	22%
Concerns with technical issues (i.e., scaling, hallucinations, ease of maintenance)	21	20%
The technology is moving too fast for credible assessment of value	18	17%
Access to specialized software	17	17%
Lack of demonstrated return on investment	17	17%
High/uncertain operational costs	15	15%
Level of AI vendor or 3rd party expertise	13	13%
Long/uncertain implementation times	12	12%
Confusion/uncertainty with vendor selection	4	4%
None of these are barriers for us	4	4%
Other	5	5%

# Coming Next: The ROI With AI Study

*Focused on return on investment, management of challenges, and shifting allocation of resources*

## Example questions:

- **To what extent did integrating generative AI models into your HPC workload environment meet your performance and cost expectations?**
- **How have budgetary plans to support gen-AI changed over the last 12-18 months?**
  - Major growth!
- **If there have been measurable monetary gains from HPC/AI integration, how long will it be to recover your initial investment?**
- **To what degree will your organization expand or contract gen-AI development moving forward?**

# In Summary

# Conclusions

- **2024 was a strong growth year**
  - GPUs, cloud, AI/ML/DL/LLM are high growth areas
  - QC systems are being installed around the world
  - And 2025 is growing by 25% for the first half of the year
- **New technologies are showing up large numbers:**
  - 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, power, talent and political changes**

# A Concern: AI Expertise Shortage

*The growing scarcity of HPC and AI experts to implement new technologies is the number one roadblock for many HPC sites*

- **Two major trends:**
  - 1) A shrinking HPC workforce
  - 2) A massive increase in system complexity and new technologies
- **HPC experts are an aging workforce**
  - The pipeline of new HPC staff entering the workforce does not match the outflow of retirees
  - Competition for HPC & AI staff is growing rapidly
- **Increasingly complex workloads are more difficult to manage**
  - Increasing number of AI & HPC systems per site
  - Augmenting traditional modeling/simulation with AI and big data
  - Incorporating multiple processor types, co-processors, accelerators, and other specialized hardware
  - Balancing on-prem and cloud
  - Enterprise IT users are entering HPC space and need HPC expertise
- **AI & HPC users will need major improvements in ease-of-use, ease-of-selection, and ease-of-optimization**

# We Welcome Questions, Comments and Suggestions



Please contact us at:  
[info@hyperionres.com](mailto:info@hyperionres.com)

# Next On The Agenda

## 10:00 - 10:30 Welcome/Introductions:

- Jacques-Charles Lafoucriere, Jean-Philippe Nominé, Maike Gilliot, Piyush Mehrotra, Rupak Biswas and Earl Joseph
- HPC Market Update, Earl Joseph, Hyperion Research

*(Session Leader: Piyush Mehrotra)*

## New Developments in HPC and AI, Part I:

**10:30 - 11:00 CEA Site Update**, Gilles Wiber, Deputy Head to the HPC Division at CEA-DAM

**11:00 - 11:30 Getting France Ready for Exascale Computing & AI**, Stéphane Requena, CTO-GENCI

**11:30 - 11:45 Vendor Update**, Tommaso Macri, QuEra

**11:45 - 12:15 Beyond Computing Limits: The AI Journey at CINECA**, Daniele Cesarini, Project Manager & HPC Technology Specialist at CINECA

**12:15 - 13:30 Lunch**

*(Session Leader: David Martin, Lawrence Livermore National Laboratory)*

## New Developments in HPC and AI, Part III:

**13:30 - 14:00 EuroHPC and AI Factories**, Josephine Wood, EuroHPC JU, Head of Unit Strategy and Governance

**14:00 - 14:20 The Network as a Scientific Instrument**, David Martin, Manager, Lawrence Livermore National Laboratory

**14:20 - 14:50 HPC at ECLAIRION**, Arnaud Lépinos, Group CEO, HPC Capital & Eclairion

**14:50 - 15:05 Vendor Update**, Bruno Lecointe, Eviden

**15:05 - 15:30 Coffee Break**

*(Session Leader: Jean-Philippe Nominé, CEA)*

# 2024 HPC/AI Market By Vendor

*The highest growth in over two decades (23.4%)!*

2024 HPC/AI Market By Vendor		
Vendor	2024 Server Revenues	2024 Market Shares
HPE	7,151	28.2%
Dell Technologies	3,916	15.5%
Lenovo	1,450	5.7%
Inspur	1,082	4.3%
Atos	708	2.8%
Sugon	619	2.4%
IBM	332	1.3%
Penguin	356	1.4%
Fujitsu	233	0.9%
NEC	213	0.8%
Other HPC	2,337	9.2%
Non-Traditional Suppliers	6,934	27.4%
<b>Total</b>	<b>25,332</b>	<b>100.0%</b>
<i>Source: Hyperion Research, 2025</i>		