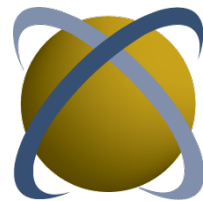


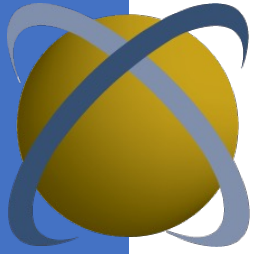
HPC-AI Market Trends 2024

SHPCP Annual Technology Meeting

April 2024



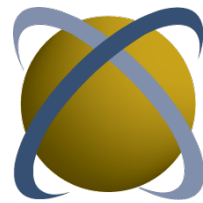
Intersect360
R E S E A R C H



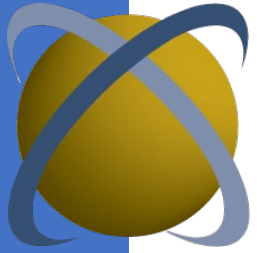
HI ket Trends 2024

SHPCP Annual Technology Meeting

April 2024



Intersect360
R E S E A R C H



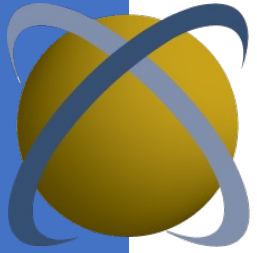
Hey Intersect360 Research ...

What is the size of the AI market?

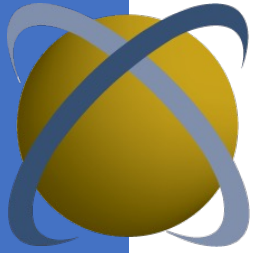


P.S. I do not have the rights for any of the images in this presentation.
I made these slides late last night, and I have COVID. Sorry.

©2024 Intersect360 Research

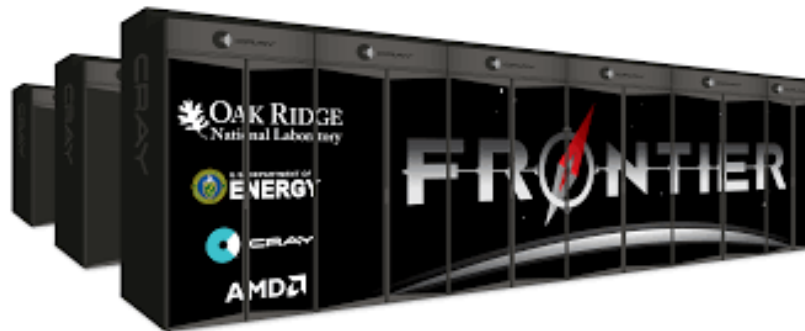


- Scalable data center infrastructure
- Intended to run AI training or inference as a primary application

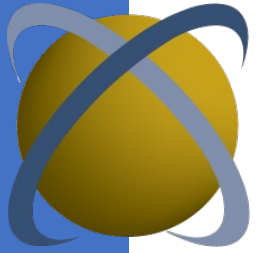


The HPC-AI Market

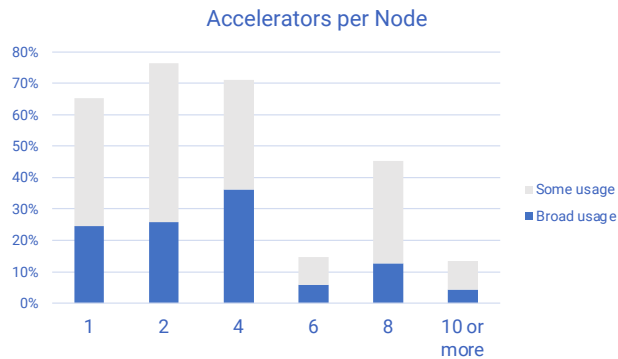
Mostly HPC, but also AI



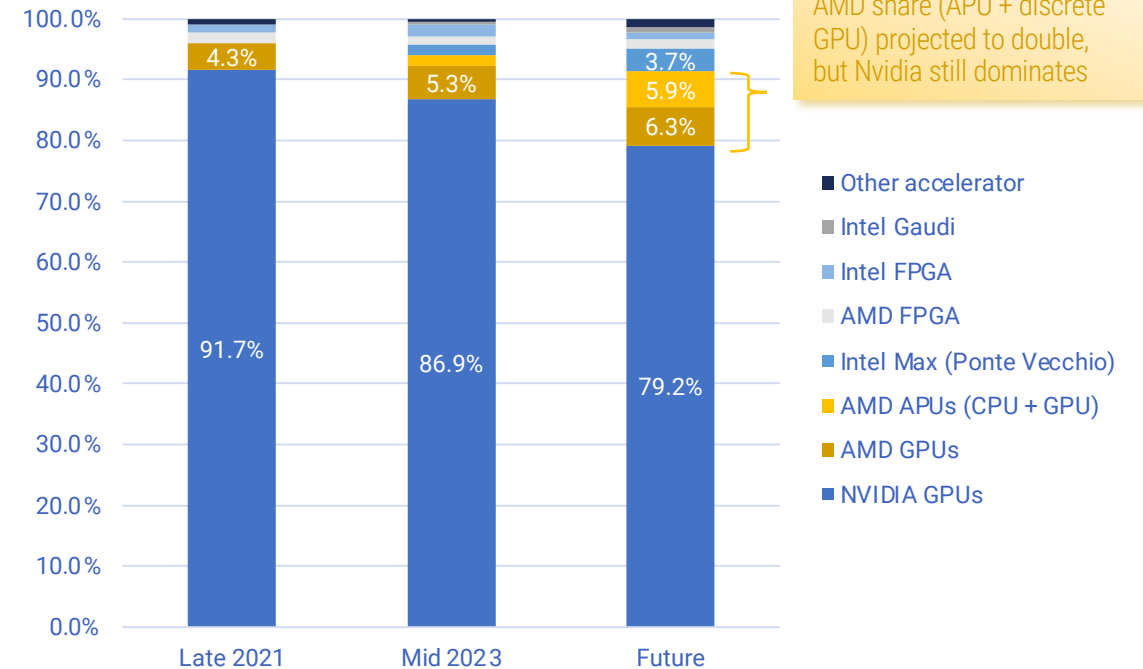
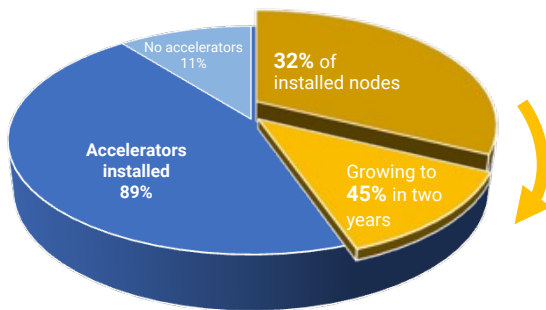
Mostly AI, but also HPC



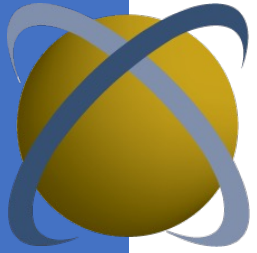
Usage of Accelerators



Four GPUs / node remains the most common configuration, "balancing" technical computing and AI



"Late 2021" represents previous survey iteration. "Future" is current survey respondents' expectation of usage in two to three years.



Accelerator Utilization

19%
say

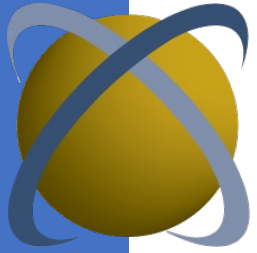
their accelerators are not highly utilized
(15% in previous survey)

30%
say

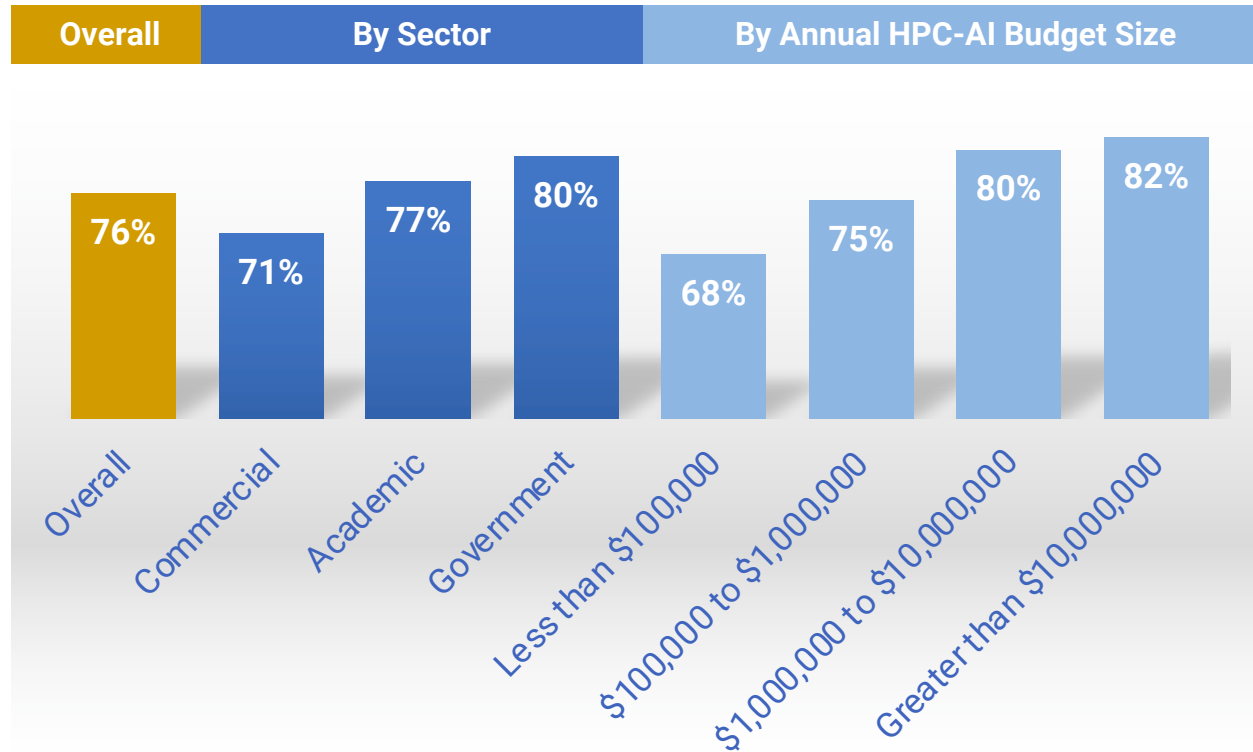
their accelerators are not as highly utilized as expected
(30% in previous survey)

25%
say

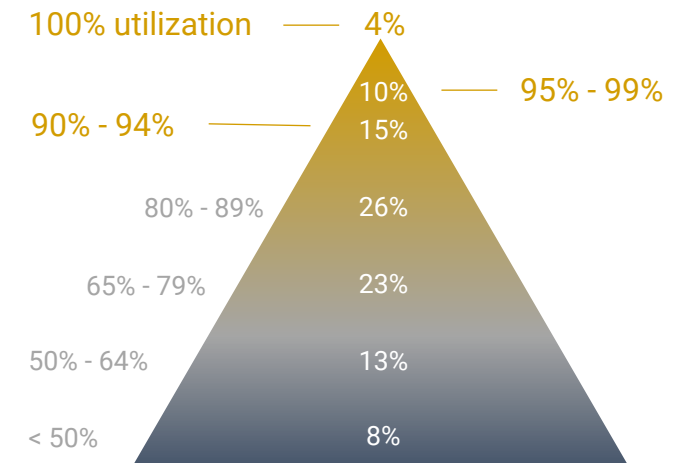
their accelerated nodes are “compute islands” not easily
accessed by other systems (33% in previous survey)

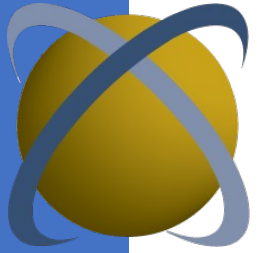


HPC-AI System Utilization



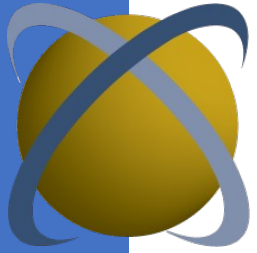
Small, commercial HPC-AI users were far more likely to have low system utilization. These are also the users most likely to use cloud.





Components of the “AI market”

- Hyperscale / Cloud: Biggest “pure AI” contribution, by far
- HPC market: Extending environments to incorporate AI
- “Enterprise AI”
 - Most of this has been in the cloud; infrastructure owned by Hyperscale
 - On-prem portion is small, but fast-growing
- Everything else: cars, smart toasters, phones, other server infrastructure ...
 - AI will be built into these things, in many ways revolutionary
 - These markets are already counted as markets for cars, toasters, etc.

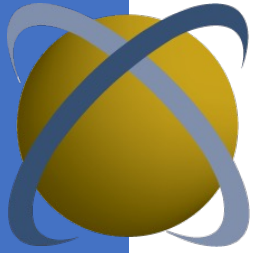


New HPC-AI Budget Research

Surveyed over 250 U.S. enterprises with at least \$100M in annual revenue

Looking for HPC or AI (or both) as proportion of IT spending

Combined with annual HPC-AI Budget Map survey to inform new market forecast

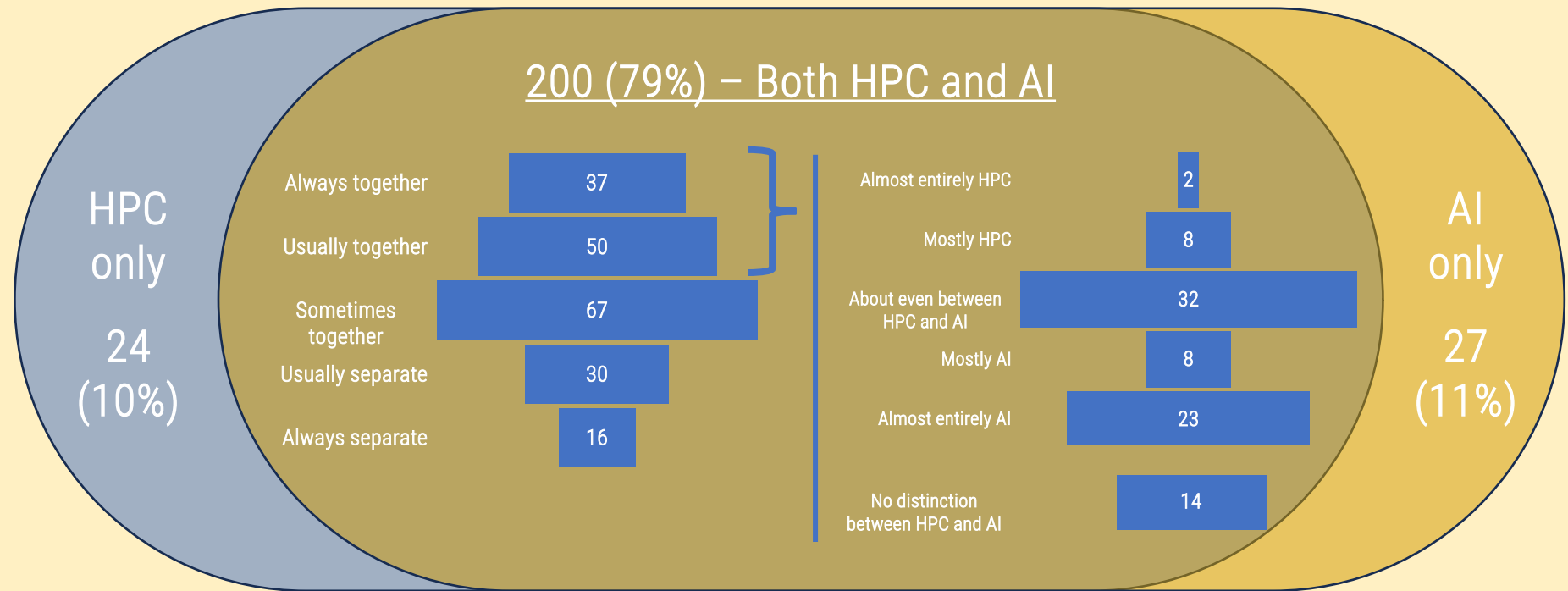


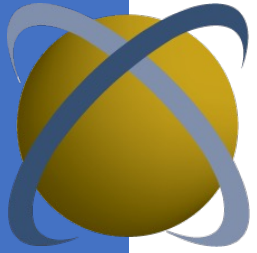
HPC-AI Budgets among U.S. Enterprise

264 total contacted

251 (95%) had HPC or AI budgets

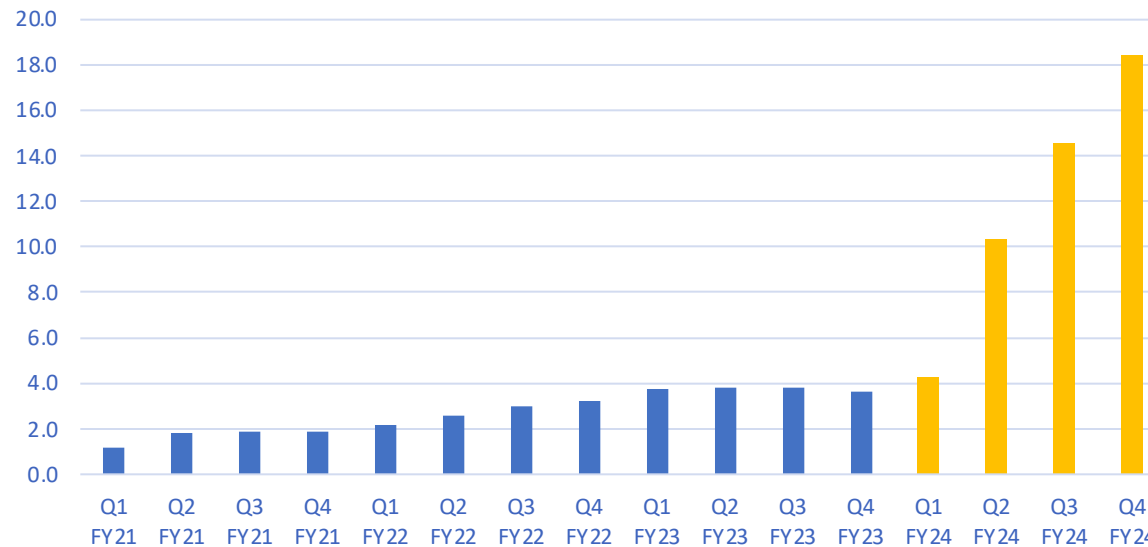
13 (5%) neither HPC nor AI





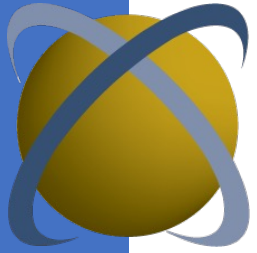
So What Do We Make of This?

Nvidia Data Center Revenue (\$B)

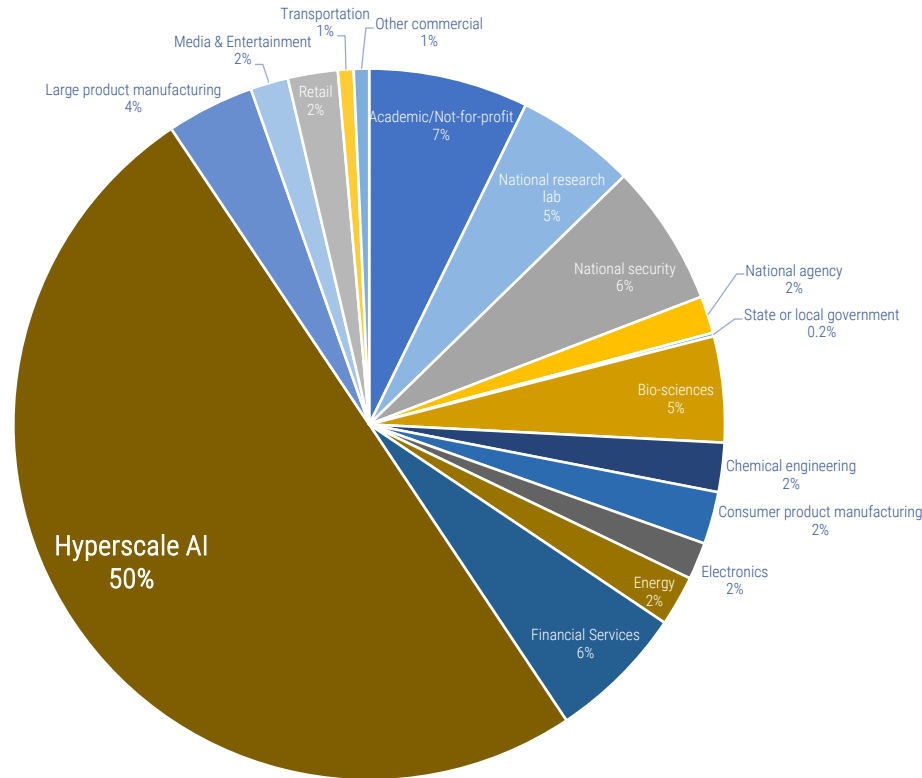


\$47.5B in FY 2024
(Feb 2023 – Jan 2024)

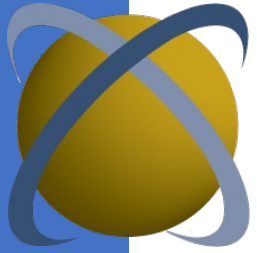
217% growth Y/Y



The Merged HPC-AI Market

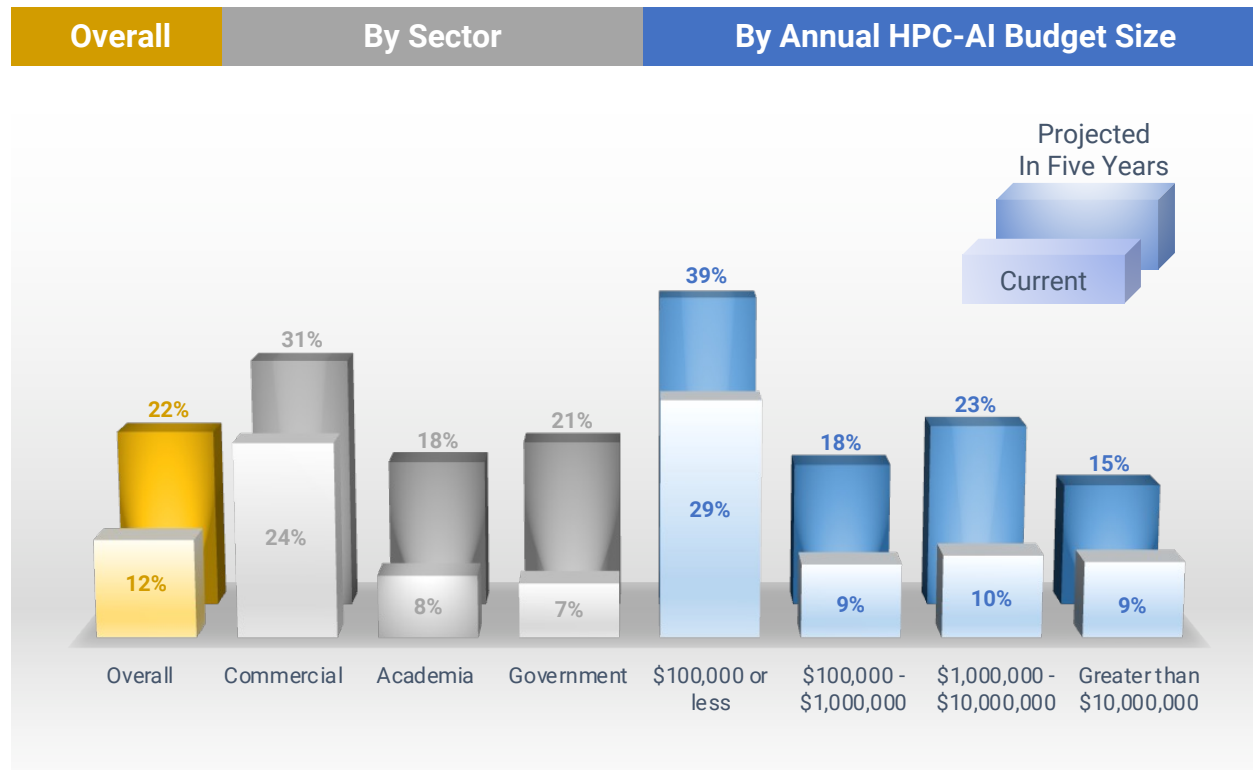


Hyperscale AI spending equaled the entirety of all other HPC-AI spending in 2023



Cloud Usage for HPC-AI

Average Percent of HPC-AI Workloads in Public Cloud



55% of HPC-AI users
use cloud at least sometimes
(Projected 77% in five years)

Cloud usage skews strongly toward smaller, commercial HPC-AI users.

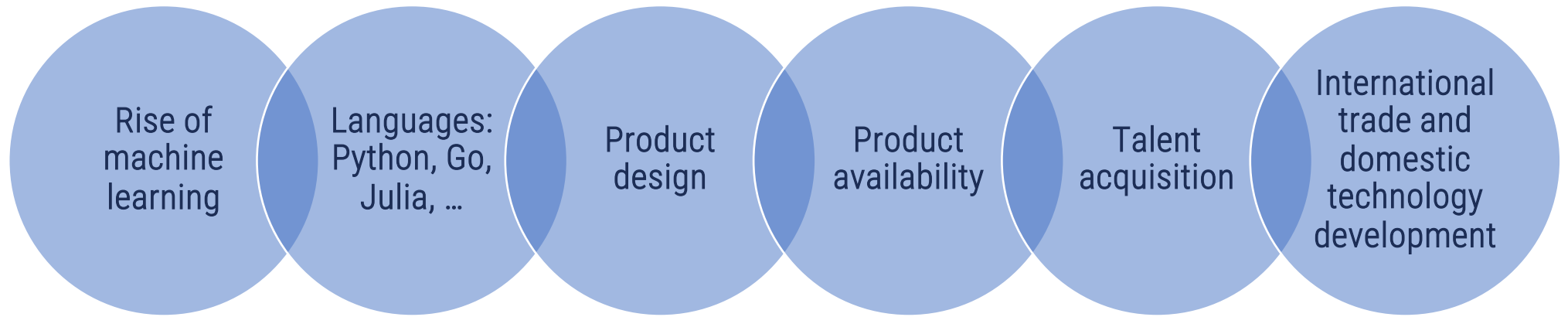
However, public sector and larger budgets are poised to post larger growth in coming years.

Cloud growth could be further accelerated by generative AI.

Cloud is now having a **significant impact** on market dynamics.



How Hyperscale is Affecting HPC





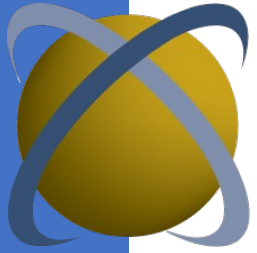
Hey Intersect360 Research ...

Have you ever Seen
anything with as much
hype and promise to
change the world as AI?



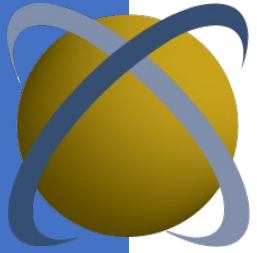
WWW .



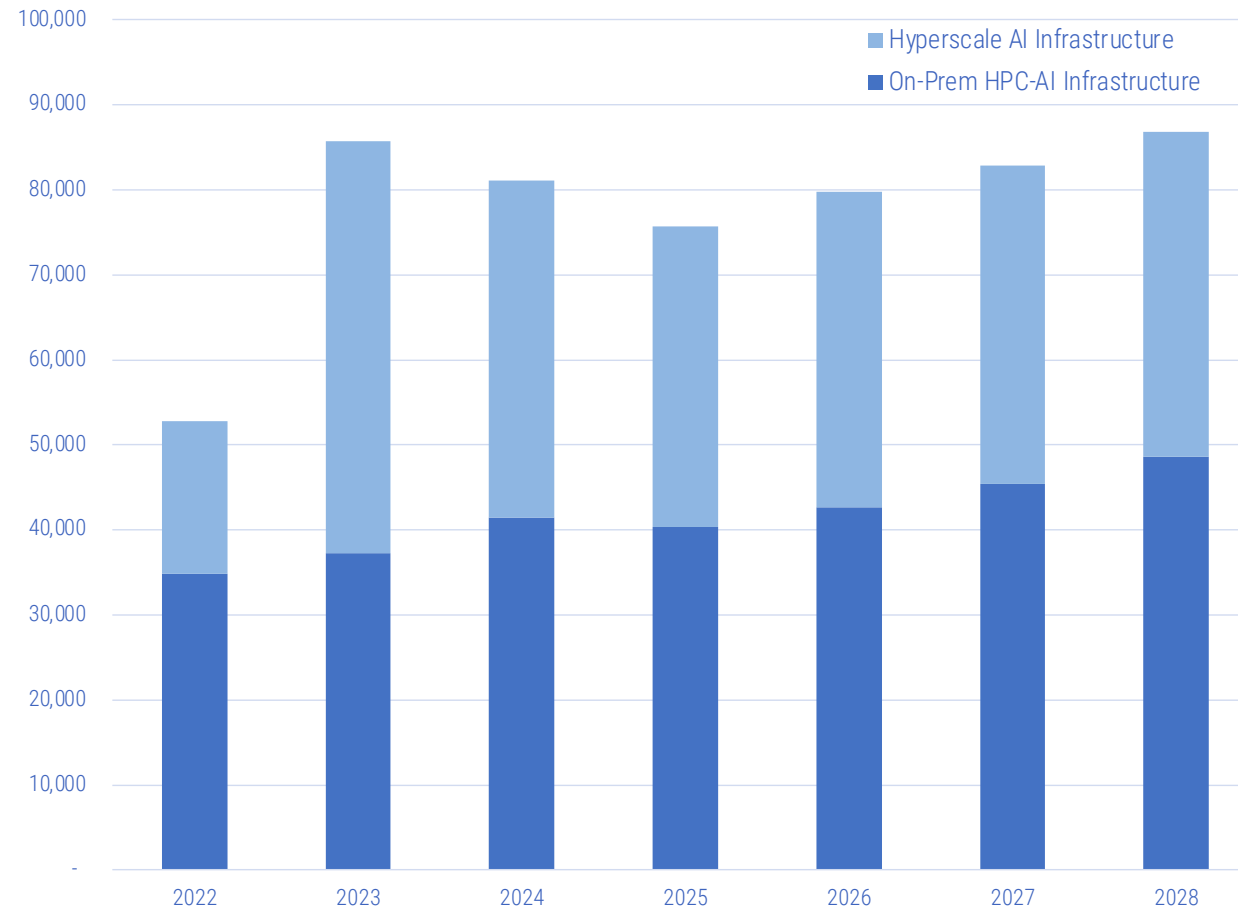


“Common Knowledge”

1. Growth in AI continues unbounded. ... *How much can the hyperscale companies absorb? This is a high-water mark in AI penetration. Big guys have been doing layoffs. Was 2023 a peak year in spending?*
2. Growth in Cloud continues unbounded. ... *Data points to no. The next five years should see a slowing in cloud growth with increasing on-premises.*
3. AI continues to require large quantities of powerful GPUs. ... *For special cases, yes, but we already see trends toward less computationally intense AI, and more emphasis on inference than training. In many cases, bad AI might be good enough.*



HPC-AI Infrastructure Forecast

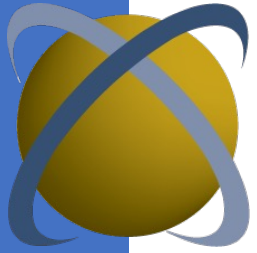


Total Market
0.3% CAGR

Hyperscale AI
-4.6% CAGR

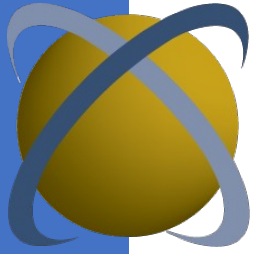
HPC-AI
5.4% CAGR

Forecast Webinar
May 2, 9:00 a.m. PT



What to Watch For

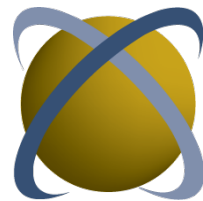
- HPC-AI Market Forecast, Including Hyperscale AI
 - Published report and press release imminent
 - Market forecast webinar, Thursday, May 2, 9:00 a.m. PT
- Highlights from HPC-AI software survey
- Intersect360 Research booth at ISC
 - 15-year anniversary of Intersect360 Research
 - Watch for a major announcement on May 2 webinar



HPC-AI Market Trends 2024

SHPCP Annual Technology Meeting

April 2024



Intersect360
R E S E A R C H