



# OrionX

## AI/ML/DL Customer Survey

July 20, 2017

Shahin Khan, Dan Olds

Orion<sup>+</sup>X<sub>.net</sub>

# OrionX AI/ML/DL Survey Goals

## + Current approaches and future plans regarding:

- Artificial Intelligence (AI)
- Machine Learning (ML)
- Deep Learning (DL)

For example:

- Budgeting
- Current status
- Technology choices
- Decision making
- Vendor selection criteria
- etc.

## + Envisioned uses of survey results:

- Help set strategy
- Help guide investments
- Validate assumptions based on industry pulse and customer sentiment
- Identify existing strengths of offerings, and areas of opportunity or concern
- Customer perspective to help position future offerings
- Credible and compelling content to fill gaps between product launches

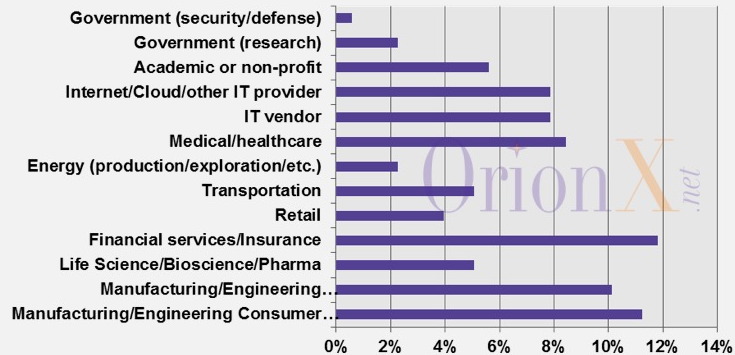
# Survey Logistics / Statistical Summary

- ✦ Survey was produced in Q2-Q3 of 2017
- ✦ Target respondents: those with expertise in and involvement with AI/ML/DL projects
  - Validated via initial questions and some level of cross-references on answers
- ✦ Total respondents: 308 individuals in 308 organizations
  - Survey ran for a period of 4 weeks.
  - Results were tabulated every 3 days after the 3<sup>rd</sup> week.
  - Survey continued to run until results remained statistically unchanged over the final 50 responses.
- ✦ Confidence interval: 95%
- ✦ Margin of error: 5.58%
- ✦ More than 140 individual questions/data points

# Respondents' Organization

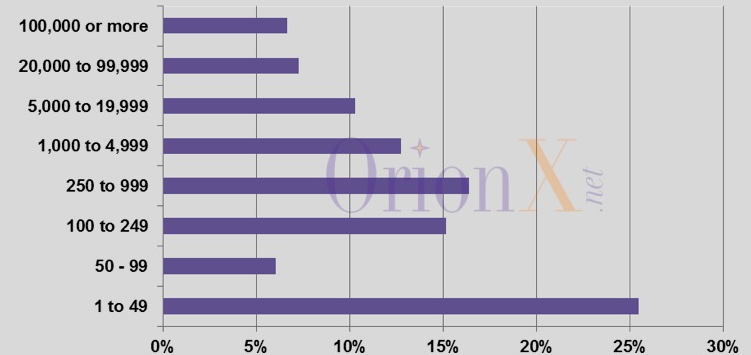
## Industry

Respondent Industry



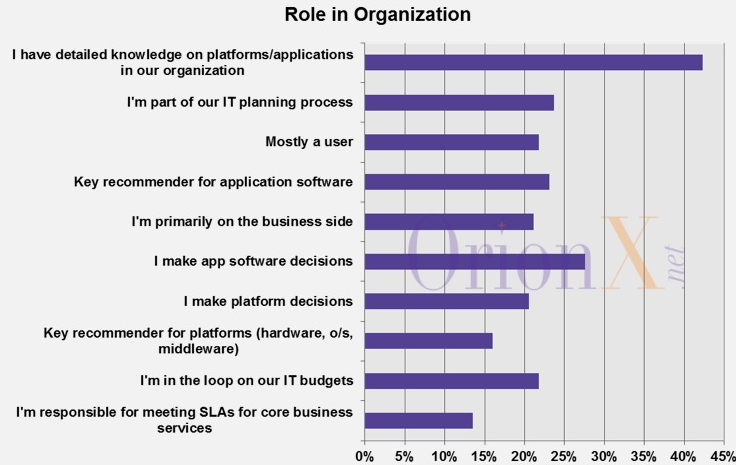
## Number of employees

Employees



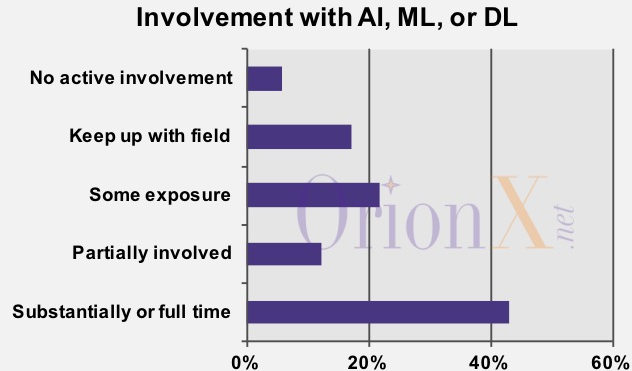
# Respondents

## Role in Organization

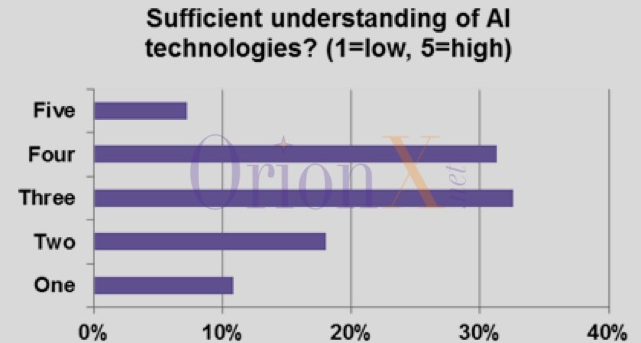


# Respondents

## Current Involvement with AI



## Understanding of AI



# Taxonomy

## DW / BI / BigData / Analytics?

Do you see Data Warehousing, BI, and Big Data/Analytics as.....?

...Completely different  
...Milestones in our path towards AI  
...Subsets of AI  
...Supersets of AI  
0%

OrionX<sub>net</sub>

## DL / ML / AI ?

Do you see a difference between the terms AI, Machine Learning, and Deep Learning?

No, they're all the same bucket

Yes, they refer to very different things

OrionX<sub>net</sub>

## What label do you use?

What label do you use to refer to the main focus of your internal projects?

Deep Learning (DL)

Machine Learning (ML)

Artificial Intelligence (AI)

OrionX<sub>net</sub>

# Current AI Budgets – Hardware

Current HW spending slightly favors cloud

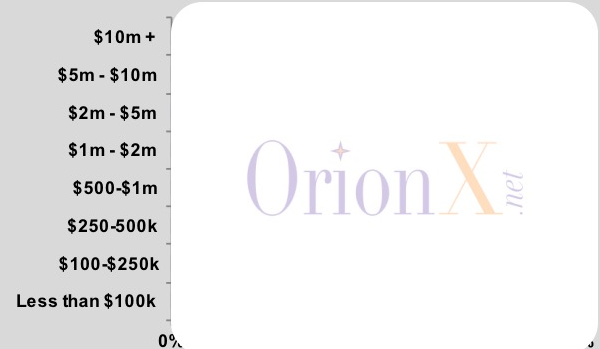
## On-premises HW

AI Budgets: On Premise Hardware



## Off-premises HW or Cloud

AI Budgets: Off Premise or Cloud HW





# Future AI Budgets – Hardware

Growth in HW spending slightly favors cloud

## On-premises HW

### AI Future Budgets: On Premise Hardware



## Off-premises HW or Cloud

### AI Future Budgets: Off Premise or Cloud HW



# Current AI Budgets – Software & Applications

Major impact by free open source software

## SW, Libraries, Frameworks

AI Budgets: SW, including Libraries and Frameworks

\$10m +  
\$5m - \$10m  
\$2m - \$5m  
\$1m - \$2m  
\$500-\$1m  
\$250-500k  
\$100-\$250k  
Less than \$100k  
0%

OrionX<sub>net</sub>

## End-user App License Fees

AI Budgets: End User App License Fees

\$10m +  
\$5m - \$10m  
\$2m - \$5m  
\$1m - \$2m  
\$500-\$1m  
\$250-500  
\$100-\$250  
Less than \$100

OrionX<sub>net</sub>

# Future AI Budgets – Software & Applications

Growth in SW spending is expected

## SW, Libraries, Frameworks

### AI Future Budgets: End User App License Fees



## End-user App License Fees

### AI Future Budgets: End User App License Fees



# Current AI Budgets – Services, Consulting, Maintenance

Significant spending on PS indicates skill shortage

## Service and Maintenance

### AI Budgets: Service & Maintenance Fees

\$10m +  
\$5m - \$10m  
\$2m - \$5m  
\$1m - \$2m  
\$500-\$1m  
\$250-500k  
\$100-\$250k  
Less than \$100k



## Professional Services / Consulting

### AI Budgets: IT Prof. Services/Consulting

\$10m +  
\$5m - \$10m  
\$2m - \$5m  
\$1m - \$2m  
\$500-\$1m  
\$250-500k  
\$100-\$250k  
Less than \$100k



# Future AI Budgets – Services, Consulting, Maintenance

Skill shortage is expected to continue, benefiting PS

## Service and Maintenance

AI Future Budgets: Service & Maintenance Fees

Up >50%  
Up 50%  
Up 25%  
Up 10%  
Same  
Down 10-50%  
0%

Orion<sup>+</sup>X<sub>.net</sub>

## Professional Services / Consulting

AI Future Budgets: IT Prof. Services/Consulting

Up >50%  
Up 50%  
Up 25%  
Up 10%  
Same  
Down 10-50%  
0%

Orion<sup>+</sup>X<sub>.net</sub>

# Current AI Budgets – Training, Staffing

## Education & Training

AI Budgets: Education & Training



## Staffing & Recruitment

AI Budgets: Staff Comp./Recruitment



# Future AI Budgets – Training, Staffing

Largest spending growth is to address skill shortage

## Education & Training

AI Future Budgets: Education & Training



## Staffing & Recruitment

AI Future Budgets: Staff Comp./Recruitment



# AI Decision Makers

Decision making is similar to other IT spend

## AI Framework

Who makes the real decisions on the AI Framework?

AI team lead

Director

CIO

OrionX<sup>+</sup>.net

## Application Architecture

Who makes the real decisions on the Application Architecture?

AI team lead

Director

CIO

OrionX<sup>+</sup>.net

## Hardware / Systems

Who makes the real decisions on the AI Hardware/Systems

AI team lead

Director

CIO

OrionX<sup>+</sup>.net



# Why / How AI?

Main expected benefits of AI are competitive advantage and cost reduction

## What department driving?

What departments are driving your AI efforts?



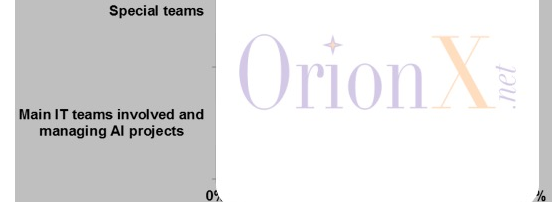
## What benefits?

What benefits do you expect from your AI efforts?



## Tiger team?

Is AI handled by normal IT or tiger team with separate budget?



# AI Visual Models

Visual recognition AI is not dominant

## Oxford VGG-16

AI Models: Oxford VGG-16

I'm an expert  
I've used it  
Planning to use  
I've heard of it  
Not familiar

OrionX<sup>+</sup>.net

## NVIDIA DetectNet

AI Models: NVIDIA DetectNet

I'm an expert  
I've used it  
Planning to use  
I've heard of it  
Not familiar

OrionX<sup>+</sup>.net

## Microsoft ResNet50

AI Models: Microsoft ResNet50

I'm an expert  
I've used it  
Planning to use  
I've heard of it  
Not familiar

OrionX<sup>+</sup>.net

# On-premises or Cloud?

## Training

Do you have a preference where AI training is run?

Don't care

Public Cloud

On-prem

OrionX<sub>net</sub>

0%

%

## Testing

Do you have a preference where AI testing is run?

Don't care

Public Cloud

On-prem

OrionX<sub>net</sub>

0%

%

## Production

Do you have a preference where AI production runs?

Don't care

Public Cloud

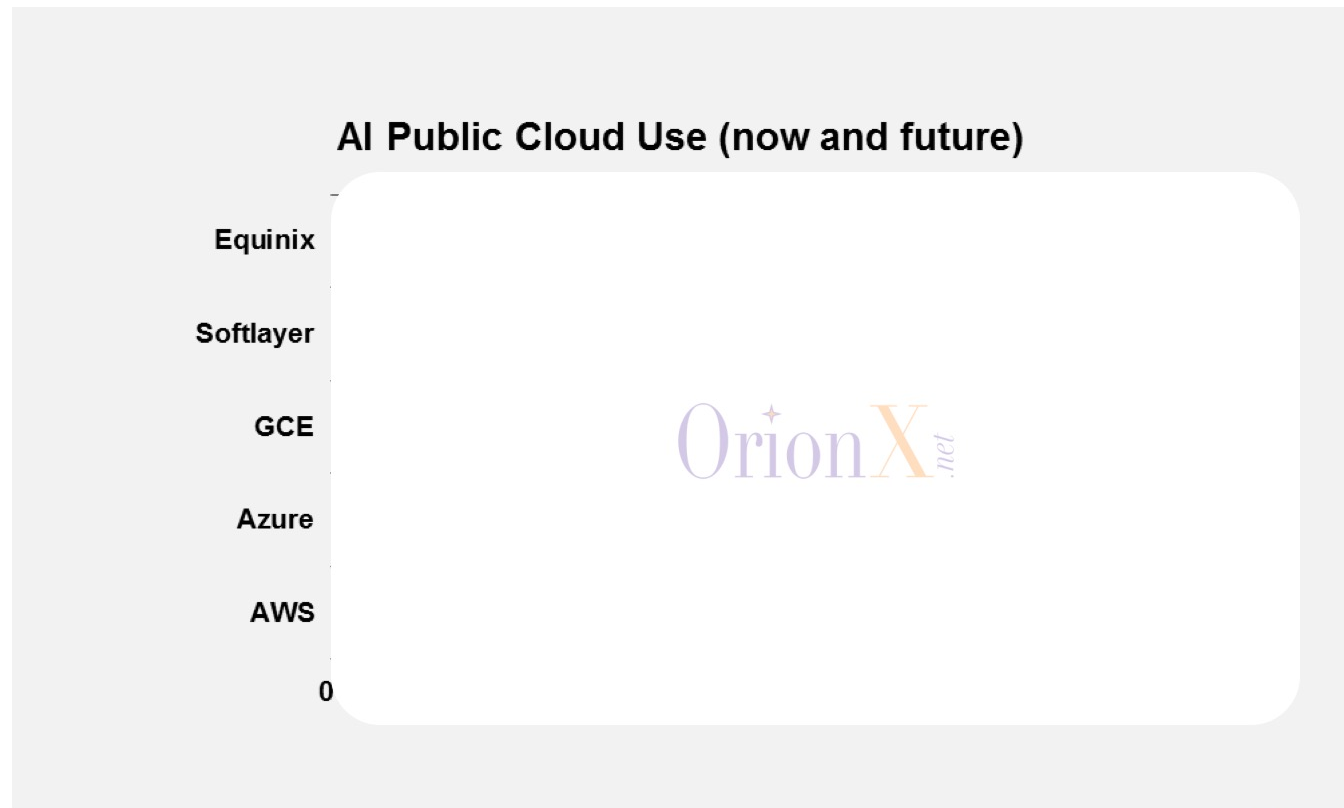
On-prem

OrionX<sub>net</sub>

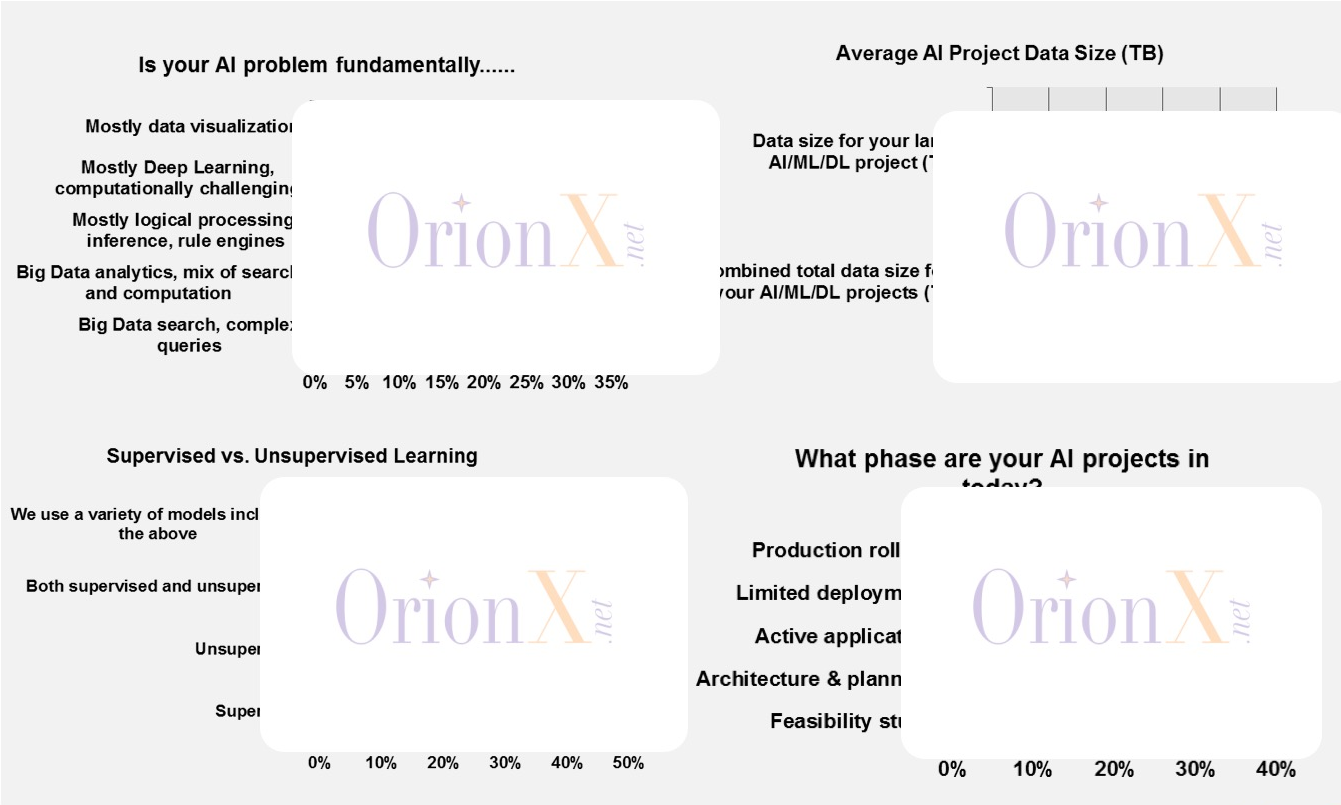
0%

%

## Public cloud preference

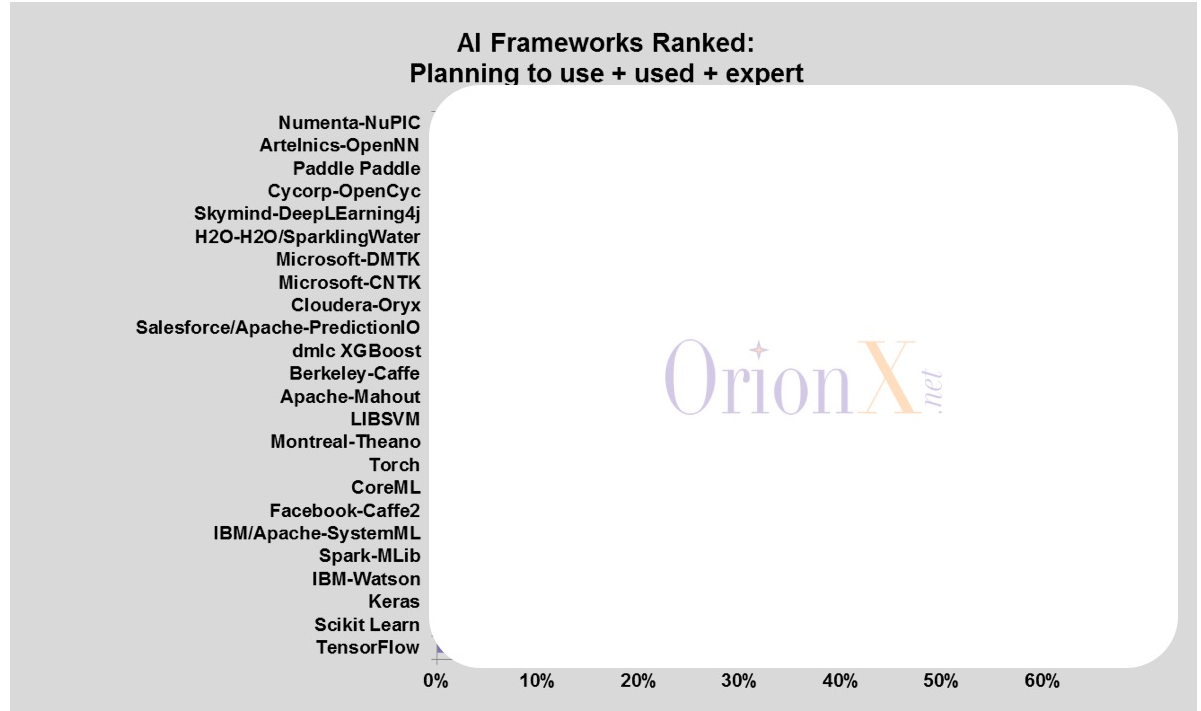


# AI Project Attributes



# AI Framework Ranking

Still an open field



# Hardware

## AI Hardware Usage (now and future)

None of these  
White box/ODM server vendor  
Oracle/Sun  
Lenovo  
IBM  
HPE  
Fujitsu  
Dell  
Cray

0.

OrionX<sup>+</sup>.net

## Hardware accelerator familiarity/usage

NVIDIA  
AMD GPUs  
Google TPU  
IBM TrueNorth  
Altera  
Mobileye  
TeraDeep  
Xylinx  
Wave Computing  
Intel Nervana  
Intel Phi  
Intel Movidius  
Graphcore  
P-Neuro  
Horizon Robotics  
BrainChip  
SpiNNaker project  
Eyeriss project  
NeuRAM3 project  
krtkl project  
KnuEdge

0.

OrionX<sup>+</sup>.net

# Data Attributes

Text and Numbers continue to dominate

## Structured?

### Structured vs. Unstructured Data

OrionX<sub>net</sub>

## Source

### AI Projects: Data Sources

Third party, other data

Sensors/IoT data

Log data

Business data, from our  
ERP system or similar

0

OrionX<sub>net</sub>

## Data type

### AI Projects: Primary Data Types

Video / Film

Audio

Images

Numbers

Text

0

%

OrionX<sub>net</sub>



# Data Quality

## Accuracy

Do you take explicit steps to ensure data accuracy?

We take some steps in filter the data but focus isn't on accuracy

No, we use what we have as is

Yes we do

OrionX<sup>+</sup>.net

## Filtering

How much data do you discard due to inaccuracy or irrelevance?

Over 75%

50-75%

30-50%

15-30%

1-15%

0%

OrionX<sup>+</sup>.net

## Reproducibility

How reproducible is your data?

Not reproducible, we have to take what we can get

We could reproduce it but with some effort

Very reproducible

OrionX<sup>+</sup>.net

# App Development

Mostly in-house, Mostly custom

## In-house or contractors / PS?

Build applications in house or hire outside contractors/consultants?

Mostly contractors

Mix of in-house and contractors

Mostly in-house

Fully In-house

OrionX<sup>+</sup>.net

## Build vs API vs Customize vs Turnkey

Propensity to build your own vs. turnkey project?

We are writing our own application from scratch including the AI libraries and framework, but will leverage what is out there

We prefer APIs that we can use to build our own custom applications

We prefer applications that we can customize

We would use a turnkey solution if it were available

OrionX<sup>+</sup>.net

## Data and Skills are the Top Challenges

## Top Challenges

### Top Challenges for AI/ML/DL Projects

Inability to leverage previous projects (those in Big Data and analytics)

Inability to leverage previous projects (those in data warehousing and business intelligence)

Getting started with the 1st project

Algorithms and models change too quickly

Quality and robustness of enabling software

Ability to hire skilled personnel

Skillset inside the organization

Organizational alignment, conflicts in objective and resources

Gathering the necessary relevant and validated data (garbage-in-garbage-out challenge)

Gathering the necessary volume of data

0%

OrionX<sub>net</sub>

# App Size and Scale

## Size

Would you characterize your AI projects as  
"large scale"?

No, they are not very large

Mid-size

Yes, they are large but will stay  
about the same size

Yes, very large and getting  
bigger

OrionX<sub>.net</sub>

## Scale

Is scaling your AI projects a challenge"?

Yes, we need more training for our  
programmers

Yes, we need better dev environment to help  
write scalable code

Yes, we need better algorithms

Yes, we need much larger memories/node

Yes, we need faster storage

Yes , we need faster interconnects

No, but our problems are too small to scale  
more than we already do

No, we can scale easily and efficiently

OrionX<sub>.net</sub>

# Cluster Configurations

## Total # nodes in all clusters

Combined number of nodes in all AI clusters

>2048 nodes  
1024-2048 nodes  
512-1024 nodes  
128-512 nodes  
32-128 nodes  
8-32 nodes

OrionX<sub>net</sub>

## # nodes in largest cluster

Number of nodes in largest AI cluster

>2048 nodes  
1024-2048 nodes  
512-1024 nodes  
128-512 nodes  
32-128 nodes  
8-32 nodes

OrionX<sub>net</sub>

## Interconnect

What cluster interconnects do you primarily use for your AI projects?

Intel OPA

InfiniBand

Whatever Ethernet standard we happen to have is fine

The fastest Ethernet we can get

OrionX<sub>net</sub>

# Cluster Node Configurations

## Memory per node

### Memory per node

128GB or more

96-127GB

64-95GB

32-63GB

16-31GB

8-15GB

OrionX<sub>net</sub>

## Accelerators per node

### Accelerators per node

3 or more GPUs or  
accelerators/node

2 GPUs or accelerators/node

1 GPU or accelerator/node

OrionX<sub>net</sub>

# Vendor Selection

## Grasp of market offerings

**Do you feel you have a good grasp of what the market has to offer for AI hardware and software**

Orion<sup>+</sup>X<sub>.net</sub>

## Product-first vs. Vendor-first

**Do you pick the product first or the vendor first?**

Orion<sup>+</sup>X<sub>.net</sub>

# Selection Criteria: Key Requirements, 1 / 2

## Skills to deploy

Key Requirements: We have the skills and expertise to deploy the product

#5  
#4  
#3  
#2  
#1

Orion<sup>★</sup>X<sub>.net</sub>

## Team is familiar

Key Requirements: What our team is familiar with

#5  
#4  
#3  
#2  
#1

Orion<sup>★</sup>X<sub>.net</sub>

## Right feature / function

Key Requirements: Right feature/function right away to solve problem (1=highest priority)

#5  
#4  
#3  
#2  
#1

Orion<sup>★</sup>X<sub>.net</sub>



# Selection Criteria: Key Requirements, 2/2

## Open Source?

Key Requirements: Open Source product

#5  
#4  
#3  
#2  
#1

OrionX<sup>+</sup>.net

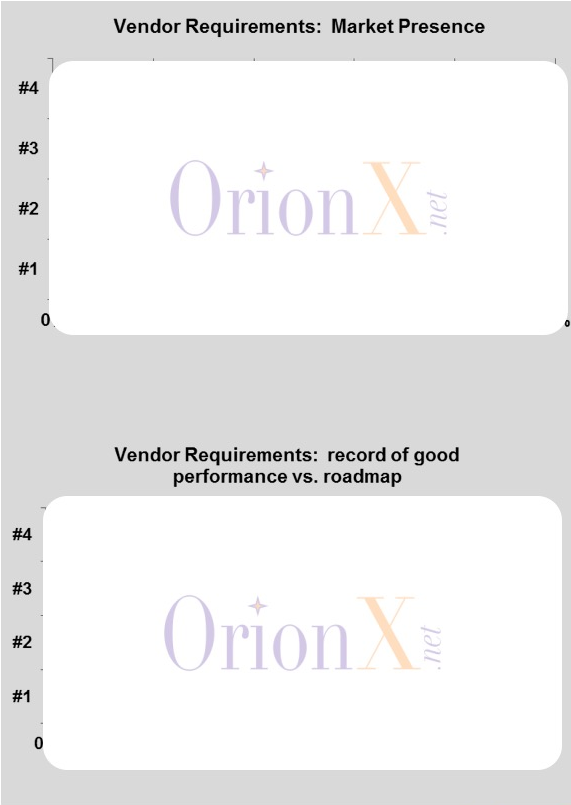
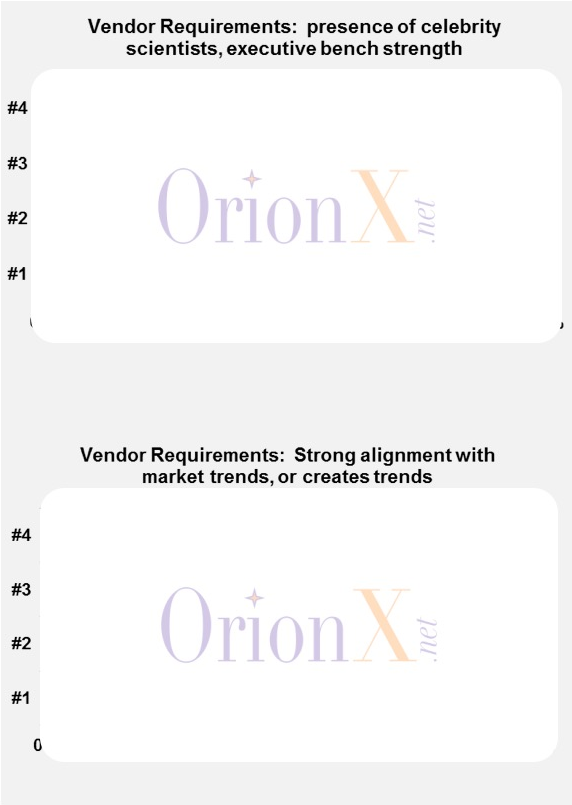
## Product Roadmap

Key Requirements: Strong product roadmap

#5  
#4  
#3  
#2  
#1

OrionX<sup>+</sup>.net

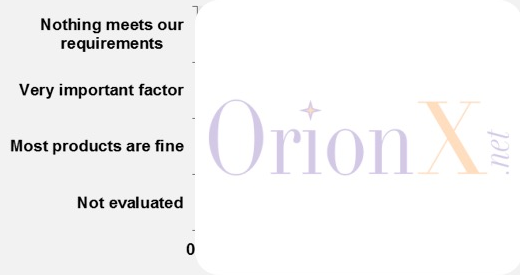
# Selection Criteria: Vendor Requirements



# Desired Solution Attributes, 1 / 2

## Performance

### Solution Attributes: Performance



## Security

### Solution Attributes: Security



## Robustness

### Solution Attributes: Robustness



# Desired Solution Attributes, 2/2

## Ease of use

Solution Attributes: Ease of use, UI/UX

Nothing meets our requirements  
Very important factor  
Most products are fine  
Not evaluated

OrionX<sup>+</sup>.net

## Ease of adoption

Solution Attributes: Ease of adoption

Nothing meets our requirements  
Very important factor  
Most products are fine  
Not evaluated

OrionX<sup>+</sup>.net

## Interoperability

Solution Attributes: Interoperability

Nothing meets our requirements  
Very important factor  
Most products are fine  
Not evaluated

OrionX<sup>+</sup>.net



Thank you!

Orion<sup>+</sup>X<sub>.net</sub>