Cyber AI Profile Working Sessions: Conducting AI-enabled Cyber Defense

August 19, 2025

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

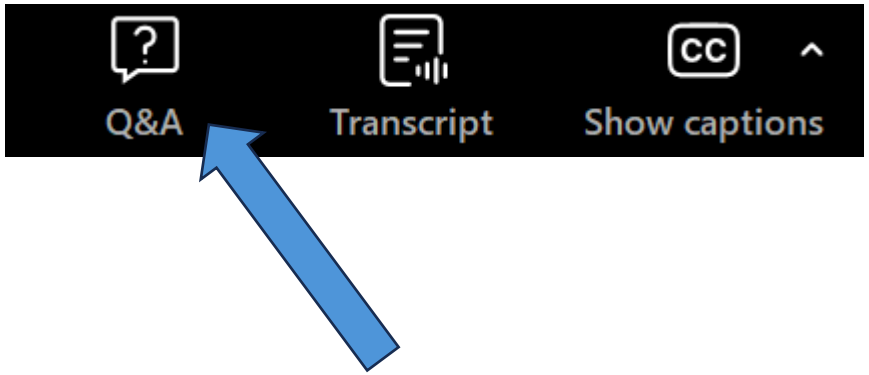
Agenda

- Cyber AI Profile Project Overview
- Today's Plan
- Refresher from Working Session Introduction
- CSF 2.0 Category Considerations: Conducting AI-enabled Cyber Defense
- Close-out

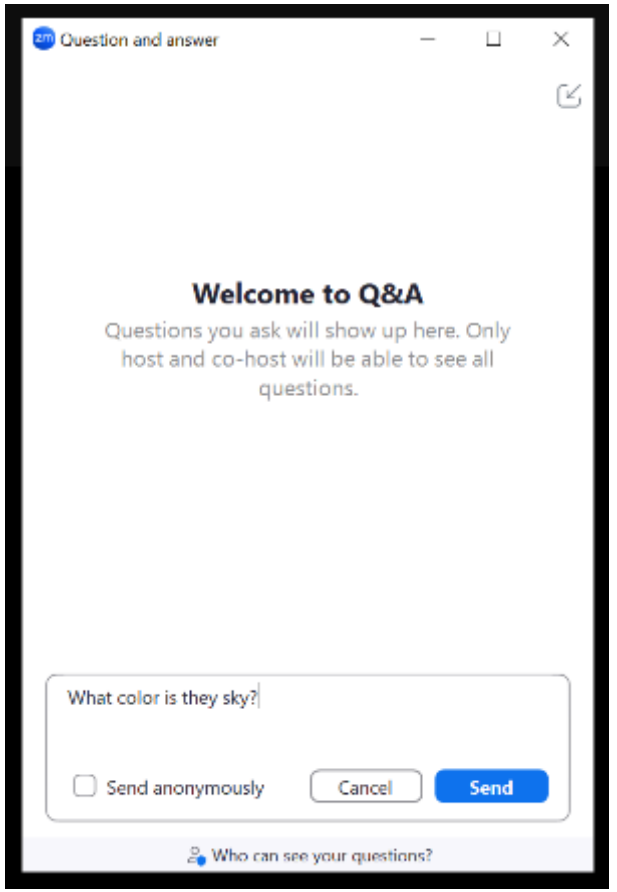
Submitting Questions

Please use the Q&A function to enter your questions.

We will do our best to answer all questions during the Q&A portion of this event.



1. To open the Q&A function, click on the “Q&A” icon at the bottom of your screen

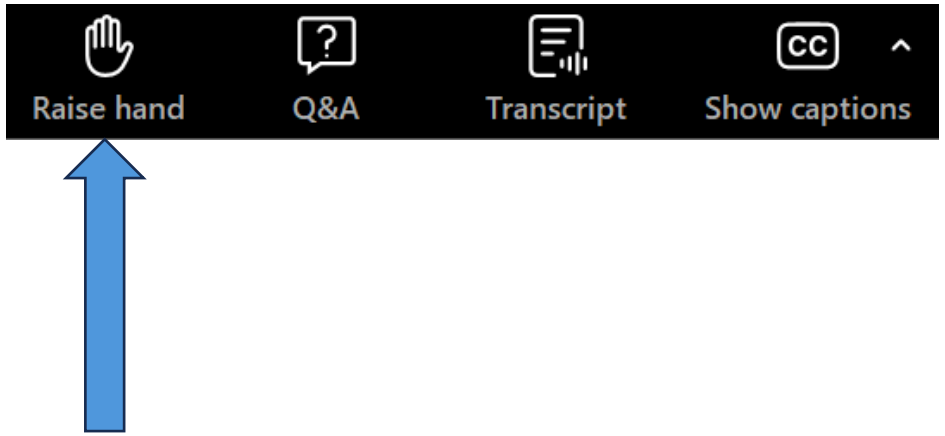


2. Type your question in the text box and click Send

Captions and Raising Hand

To enable captioning during the event, click on the “Show captions” icon at the bottom of your screen.

To raise your hand during the Q&A sessions, click on the “Raise hand” icon at the bottom of your screen.



Cyber AI Profile Project Overview

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Cybersecurity, Privacy, and AI

The diverse use and rapid proliferation of Artificial Intelligence (AI) promises unique value for industry, consumers, and broader society, but like many technologies, to recognize these benefits to the greatest potential, new risks from these advancements in AI must be managed.

In NIST's Applied Cybersecurity Division (ACD), our key concern is how advancements in the broad adoption of AI may impact current cybersecurity and privacy risks and risk management approaches.

https://www.nist.gov/itl/applied-cybersecurity/cybersecurity-privacy-and-ai

NIST Cybersecurity-focused work on AI

- AI Risk Management Framework - a framework to better manage risks to individuals, organizations, and society associated with artificial intelligence
- The Secure Software Development Practices for Generative AI and Dual-Use Foundation Models
- NIST AI 100-2 E2023: Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations
- Dioptra – a software test platform for assessing the trustworthy characteristics of artificial intelligence
- Federated Learning on Privacy Enhancing Technology (PET) - Evaluating Differential Privacy Guarantees
- TrojAI Challenge Rounds Based on Data Poisoning: Test & Evaluation of Trojan detectors
- *NIST SP 800-53 Overlays: Agents; LLM; Prediction; Classification*
- *Automotive Cybersecurity Community of Interest (COI): Community of interest examining challenges from increased cybersecurity risk and the adoption of AI and opportunities*
- National Cybersecurity Center of Excellence exploring new projects for Cybersecurity in AI and cybersecurity of AI: AI SecDevOps; Agent Identities

The Case for a Cyber AI Profile

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Purpose:

Support cybersecurity programs as they manage the impacts of advancements in AI to their organization

Areas of focus:

- Cybersecurity risks that arise from the use of AI by organizations, including securing AI systems, components, and machine learning infrastructures, and minimizing data leakage.
- Determining how to defend against AI-enabled attacks.
- Assisting organizations in the use of AI with their cyber defense activities and using AI to improve privacy protections.

Outcomes:

- Establishes a shared understanding of AI-related cybersecurity priorities and considerations for any organization
- Fosters collaboration and communication across the AI and cybersecurity communities
- Enables organizations that are using AI technologies to demonstrate a degree of commitment and trustworthiness using a common set of outcomes in the Profile

Cyber AI Profile Roadmap

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Initial Research and Scope Shaping
- Initial stakeholder discussions
- Literature review

NIST collaboration throughout

Literature review continues as needed

Blog Post / Announcement (F)
- Share intent
- Establish Community of Interest (COI)

Cyber AI Profile Concept Paper (F)
- Approach
- Format examples
- AI References
- Open invite for public to provide input

Public Workshop #1 (F)
- Concept Paper feedback

COI Working Sessions (F)

Cyber AI Profile Preliminary Draft (F)

Public Workshop #2 (F)
- Summary of feedback
- Additional inputs

Additional COI Working Sessions?

Cyber AI Profile Initial Public Draft (IPD) (F)

Public Workshop #3 (F)
- Summary of feedback
- Additional inputs

2PD?

Cyber AI Profile FINAL

Organizational Use + Additional Guidance (TBD)

F Opportunities for COI/public stakeholder feedback (NOTE: Internal NIST collaboration occurs throughout)

Today's Plan

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Working Sessions by Focus Area

Conducting AI-enabled Cyber Defense

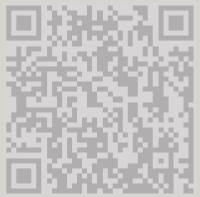
August 19, 2025

1:00–4:00pm EDT

Securing AI System Components

August 26, 2025

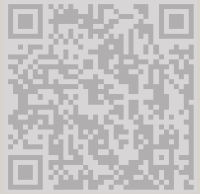
1:00–4:00pm EDT



Thwarting AI-enabled Cyber Attacks

September 2, 2025

1:00–4:00pm EDT

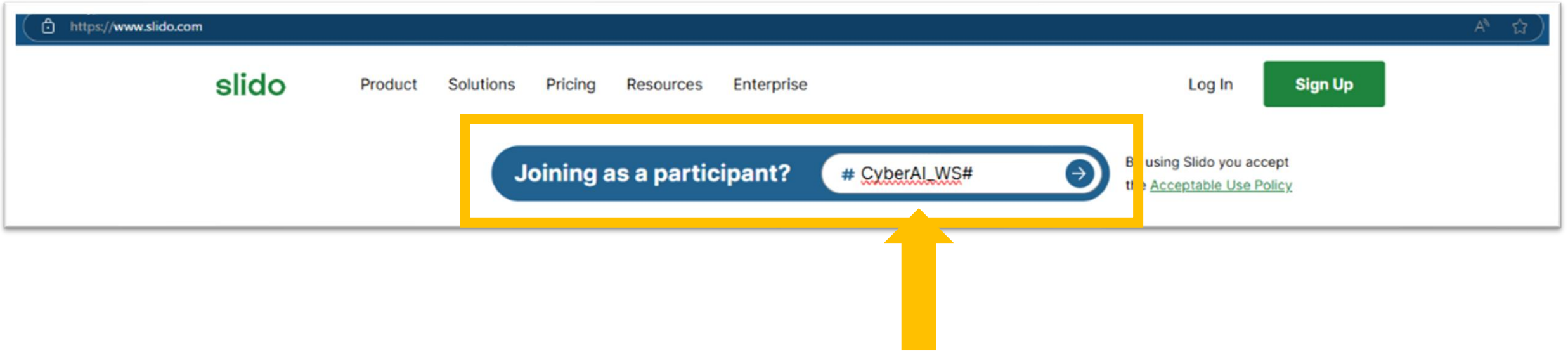


Housekeeping

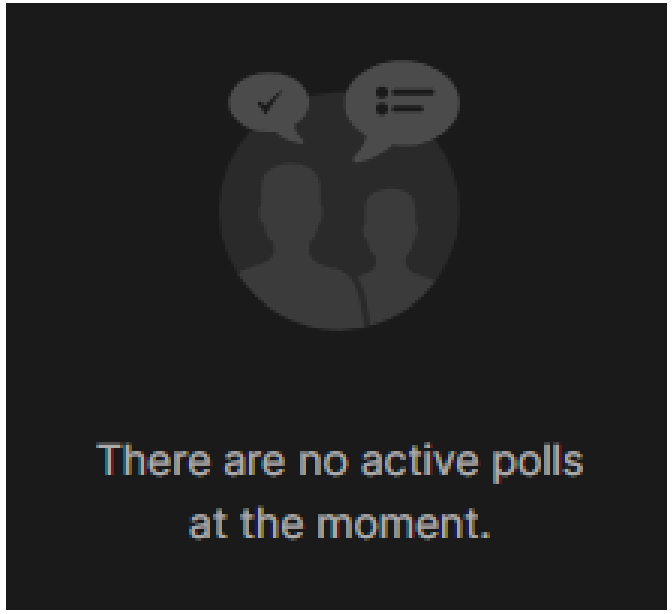
- Members of the press, please identify yourself and your organization
- We will use Slido to facilitate some of our discussions
- Please raise your virtual hand or type in the chat to contribute
- Feel free to introduce yourself as you are speaking (name and organization)
- Be respectful of others
- **Please remain on mute when not speaking**

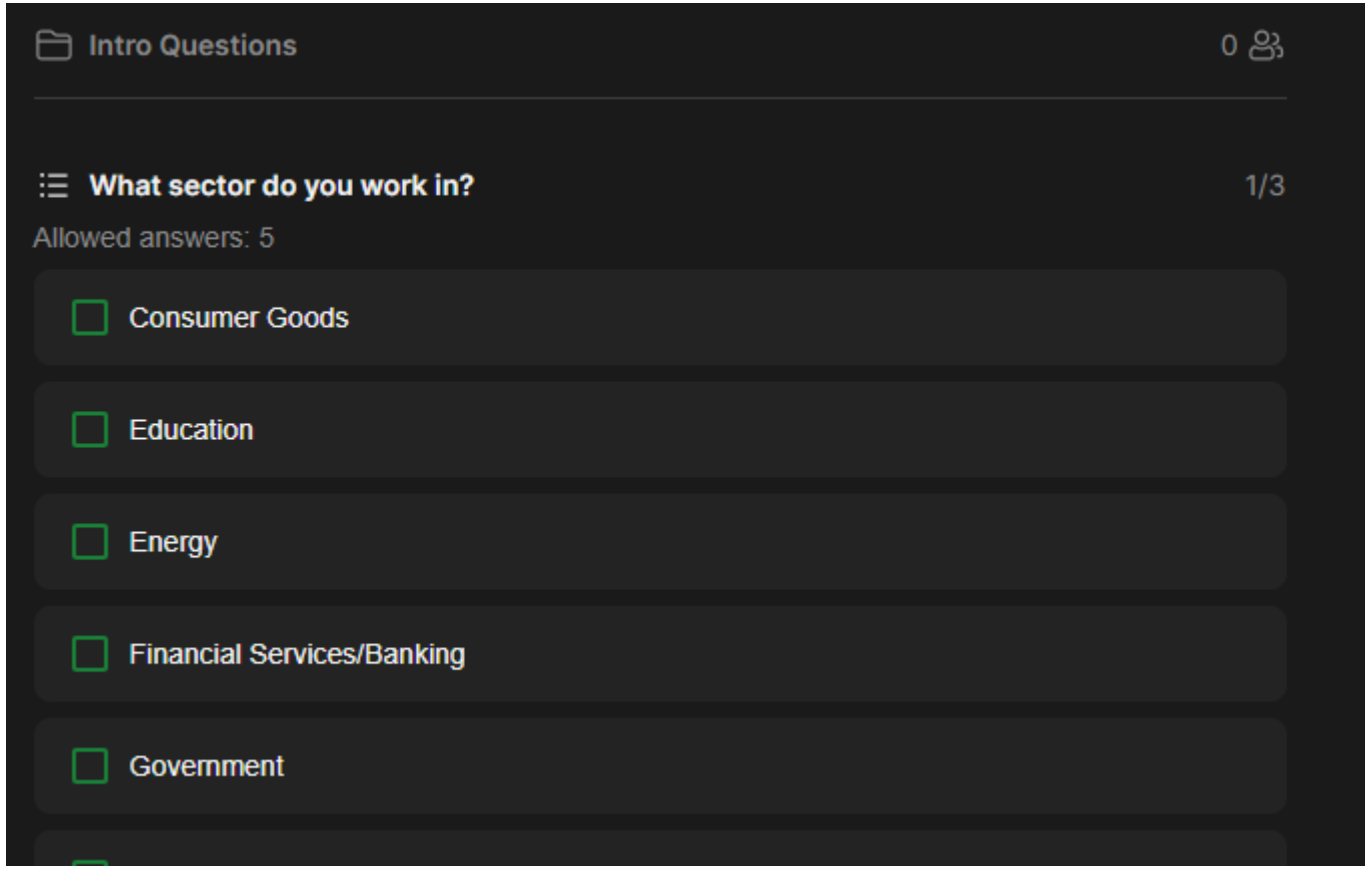
Using Slido

- We will be using Slido to facilitate some of our discussions
- Options to join via QR code or URL + event code
- Works on mobile phone and computer
- Responses are anonymous



Slido Screens

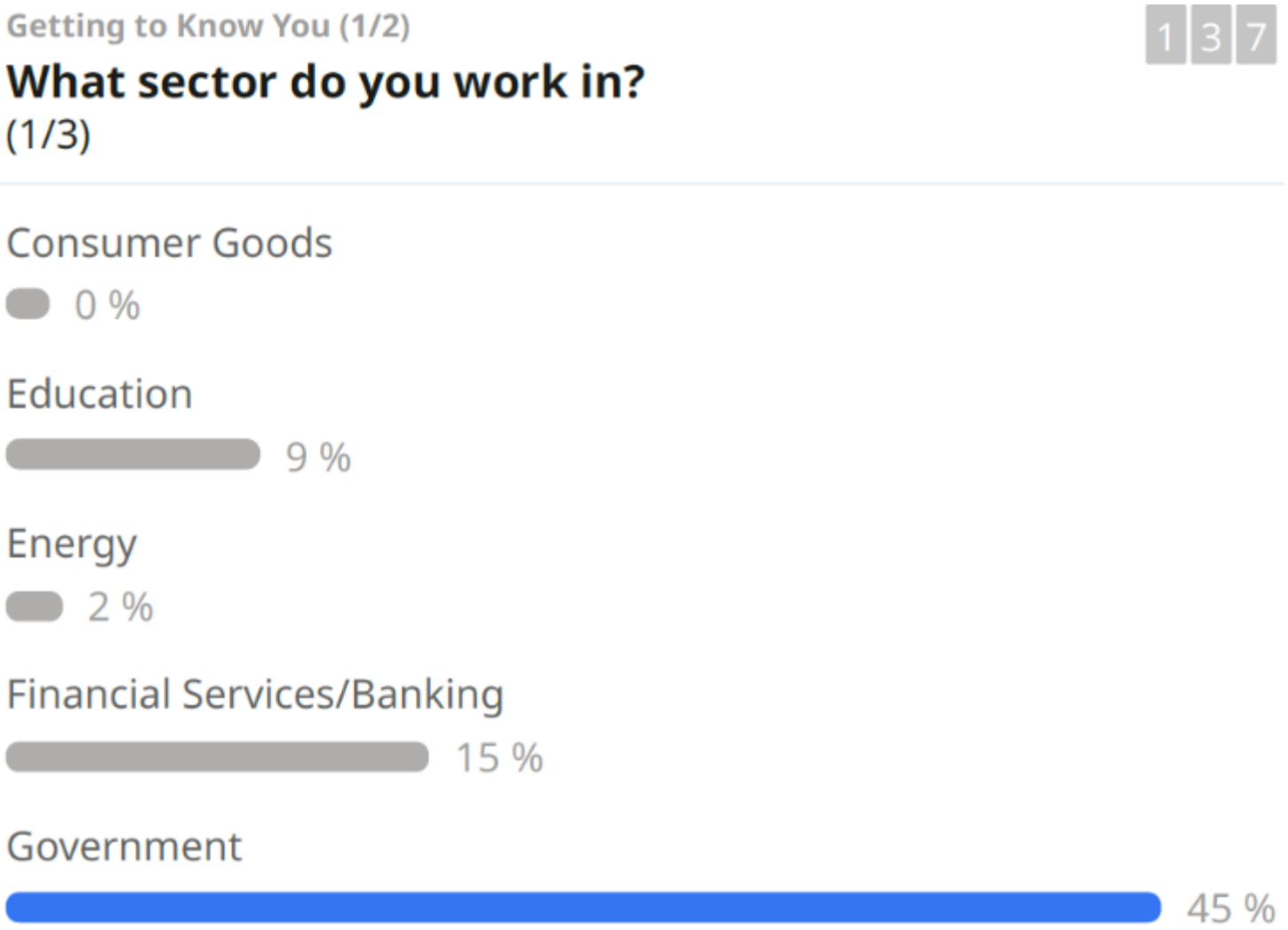




Slido: Getting to Know You

- What sector do you work in?
- Which NIST Frameworks does your organization use?

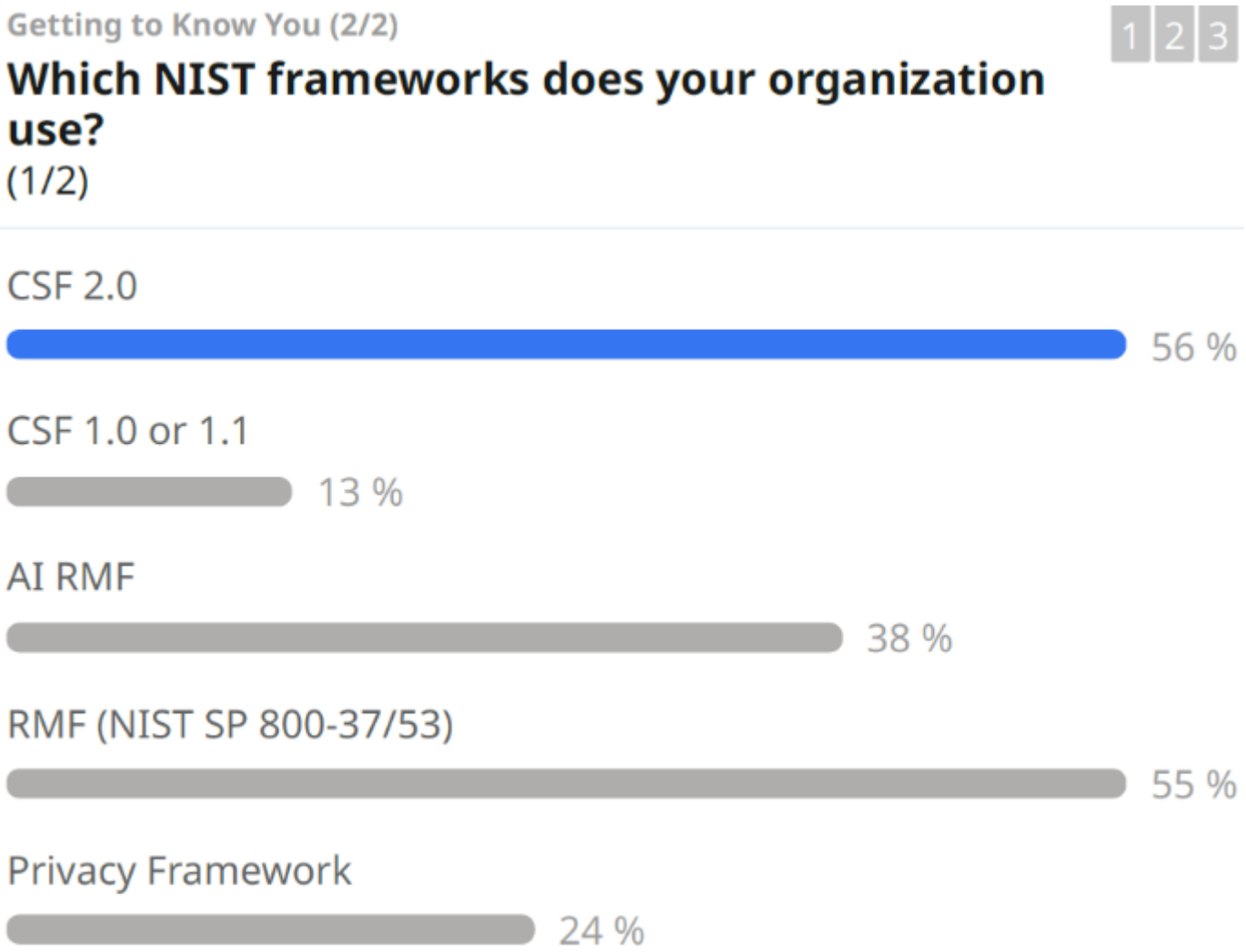
Slido Results: *Getting to Know You (1 of 5)*



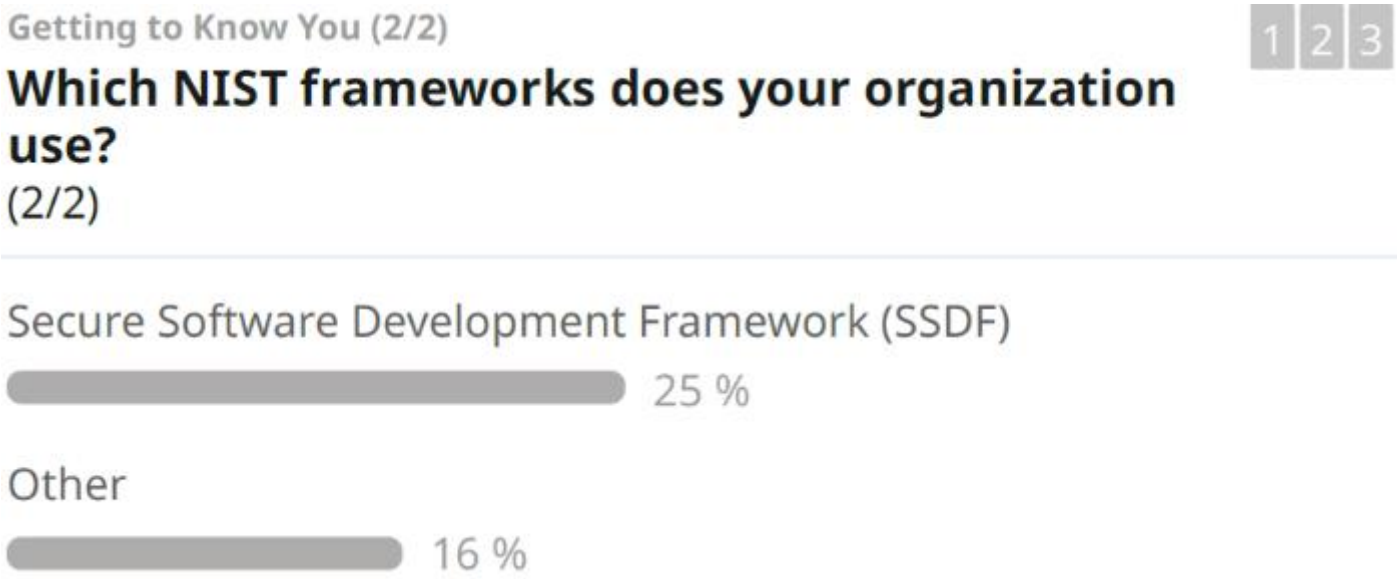
Slido Results: *Getting to Know You (2 of 5)*

Slido Results: *Getting to Know You (3 of 5)*

Slido Results: *Getting to Know You (4 of 5)*



Slido Results: *Getting to Know You (5 of 5)*



Today's Focus: CSF 2.0 Categories for Conducting AI-Enabled Cyber Defense

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

- **Scope:** Identifying opportunities for use of AI in supporting cybersecurity defense activities and understanding the challenges when leveraging AI to assist in defensive operations
- **Focus Area characteristics – cybersecurity considerations regarding:**
  - Understanding how organizations are leveraging AI to support cybersecurity defensive activities
  - Identifying challenges introduced when AI is leveraged for cybersecurity operations
    - Not functioning as expected
    - Producing false positives/negatives
    - Employee overreliance on AI outputs

Opportunities

- Advanced Threat Detection & Analysis
- Automated Incident Response
- Proactive Risk Management
- Security Governance & Policy

Challenge Areas

- Lack of Explainability
- Confabulation
- Information Integrity
- Data Privacy
- Human-AI Configuration
- Value Chain and Component Integration

Slido.com
#CyberAI_WS

Slido Results: *Conducting AI-Enabled Cyber Defense (1 of 17)*

Conducting AI-enabled Cyber Defense (1/4)

0 4 9

What additional concepts should be considered to describe this Focus Area (scope and characteristics)?
(1/7)

- Unfortunately, our agency is still in the beginning stages of AI
- LLM explainability and level of accuracy/truthfulness
- common algorithm definition for certain task
- Identifying when AI is being (perhaps silently) integrated (by employees or vendors). Education and if AI systems elsewhere are changing employee behavior.
- An additional concept is to use Cybersecurity Coaching to sus out and Identify the the Opportunities.
- controls and checks. Verifications. App and solution verbosity
- General AI security tool management. Individual companies are creating their own tools/products. We should consider focusing on how to support TRiSM, to ensure that

Slido Results: *Conducting AI-Enabled Cyber Defense (2 of 17)*

Conducting AI-enabled Cyber Defense (1/4)

049

What additional concepts should be considered to describe this Focus Area (scope and characteristics)?
(2/7)

AIs that are out there are in fact doing what we expect and monitoring the tool/product AI to look for symptoms of compromise.

- data corruption/poison esp. in backup sets
- Understanding the readiness gap ; Clearly defining the differences between AI and RPA - there is confusion at the CSuite level and too often they will invest in AI and expect RPA actions or vice versa
- Improved data integrity, better data quality assurance regulation for LLMs
- Leveraging AI for cyber defense: threat intelligence; incident response; identity governance; baselining

Slido Results: *Conducting AI-Enabled Cyber Defense (3 of 17)*

Conducting AI-enabled Cyber Defense (1/4)

049

What additional concepts should be considered to describe this Focus Area (scope and characteristics)?
(3/7)

the enterprise's baseline network ops traffic; cyber supply chain risk management

- Vulnerability Remediation Audit Log Correlation parsing out all the vulnerabilities from all the tools we use. Ransomware attack
- Potential future deviations from current trends like neurosymbolic AI
- Cloud
- None I can think of
- AI data source(s) tracking, traceability, and monitoring
- how to add enterprise risk management how to integrate NIST AI RMF into CSF 2 develop a crosswalk between NIST AI RMF, privacy, and CSF 2 how to map CSF and AI RMF to AI Action Plan
- Cyber Warfare Battlefield Management. Build an Arsenal of

Slido Results: *Conducting AI-Enabled Cyber Defense (4 of 17)*

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Conducting AI-enabled Cyber Defense (1/4)

049

What additional concepts should be considered to describe this Focus Area (scope and characteristics)?
(4/7)

Cyber Weapons as Countermeasures. A Threat Model as a Threat Landscape where Countermeasure Launch Triggers are Deterministic LLM Adaptive Reasoning Rules.

- Services to help flag vulnerabilities in Ai coding b platforms.
- insider threat
- NIST 800-63-4 (Identity)
- With AI changing at a much faster pace then even the internet did, how does these frameworks stay nimble / adapt with how AI is changing. Basically how do we keep up. Again this is around security and change
- Security workers (e.g., SOC) using chatbots to help identify if an alert is a false positive or true positive. Security workers using chatbots (eg., with RAG)

Slido Results: *Conducting AI-Enabled Cyber Defense (5 of 17)*

Conducting AI-enabled Cyber Defense (1/4)

049

What additional concepts should be considered to describe this Focus Area (scope and characteristics)?

(5/7)

to compare potential incidents to policy, procedures, etc. I would like to see equivalents to (or expanding) CVE, CWE, ATT&CK, etc. to include AI concepts.

- I really like your idea of taxonomy for ai and cyber, trying to sync them more. I always need help porting the taxonomies through the companies marketing.
- I think you have them covered. I like the thought of automated threat detection and analysis as this will support cybersecurity staff. As the other participant stated, logging analysis would be great given the voluminous amount of data we have today.
- mapping and attack path management
- Understanding what AI defense tools are available to combat an AI enabled hacking took.

Slido Results: *Conducting AI-Enabled Cyber Defense (6 of 17)*

Conducting AI-enabled Cyber Defense (1/4)

049

What additional concepts should be considered to describe this Focus Area (scope and characteristics)?
(6/7)

- Threat detection and response Network monitoring Threat intelligence integration
- Cyber threat intelligence
- AI hallucinations, and a focus on how to reduce/eliminate them.
- Using AI to locate vulnerabilities within a system
- Agentic AI security
- How ai divulging proprietary data if used by employees
- Adversarial AI (use of AI by adversaries)
- AI password cracking
- Cyberattacks prediction
- Financial Market analysts Scrubbing internal data Threat analysis Secrets in code
- Shadow AI
- Training & Education. Assist in filling the cyber resource gap, especially with college and university students.
- ZTN and AI

Slido Results: *Conducting AI-Enabled Cyber Defense (7 of 17)*

Conducting AI-enabled Cyber Defense (1/4)

049

What additional concepts should be considered to describe this Focus Area (scope and characteristics)?
(7/7)

- Forensic log analysis, counter measure recommendations
- Prebuilt AI profile per critical infrastructure industries
- Regulatory
- Phishing Attacks
- Use AI to understand what controls are in place, compliant and effective
- 1. Data/information poisoning 2. Data/information propagation - leaking
- password crackinh
- Active intrusion prevention
- none

Slido Results: *Conducting AI-Enabled Cyber Defense (8 of 17)*

Conducting AI-enabled Cyber Defense (2/4)

0 4 0

What additional types of opportunities is your organization considering for incorporating AI into cybersecurity defensive operations?
(1/4)

- 1) how to trust AI (not another attack vector) for my cyber defense program. 2) how to define responsibility model when an AI is being used in my cyber defense 3) how to define a model / vendor agnostic AI-powered cyber defense with consistent results
- Waiting for direction from leadership
- 1st Tier customer support Speech to text, text to speech Playbook creation Automation of Incident Response
- I am a Cyber security Coach.
- Analytics and for most organizations - embedded features
- I have no opinion
- incorporating a ai center-of-excellence within cyber-sec org
- Assessing devices for configuration drift in conflict with

Slido Results: *Conducting AI-Enabled Cyber Defense (9 of 17)*

Conducting AI-enabled Cyber Defense (2/4)

0 4 0

What additional types of opportunities is your organization considering for incorporating AI into cybersecurity defensive operations?
(2/4)

the accepted CSF whether it be NIST, Essential 8, or CMMC; connecting user logins on desktops/laptops to proximity of registered mobile devices;

- Constitutional Ai Governance
- Regulations and Policies Cyber Ethics for AI
- Incorporating LLMs to assist in creation of ATO packages - specifically creating policies to comply with NIST 800-53 R5 controls - Access Control Policies, Configuration Management policies, etc
- Investigation and sancioned
- None I can think of
- Tying vulnerability to potential threat actors and exploitation tactics
- how to incorporate this into ICS SCADA CBRNE environment
- Analytics and insights

Slido Results: *Conducting AI-Enabled Cyber Defense (10 of 17)*

Conducting AI-enabled Cyber Defense (2/4)

0 4 0

What additional types of opportunities is your organization considering for incorporating AI into cybersecurity defensive operations?
(3/4)

of cyber data.

- Self-Learning AI Swarms collaboration. 3 of 5 Voting Determinism of False Positive Threats. SIEM Correlation Rules as a LLM
- Identifying gaps in the the existing standards
- Its been mandated from audits to research too process improvements to identification
- AI is baked into practically all security products.
- Typically, the capabilities that are being embedded into security operations capabilities. We want to automate analysis and reporting to assist staff.
- Third party vendors
- SOC, Risk Management
- Using AI to move away from a 3rd party security providers
- Analysis of data
- Securing AI applications

Slido Results: *Conducting AI-Enabled Cyber Defense 11 of 17)*

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Conducting AI-enabled Cyber Defense (2/4)

0 4 0

What additional types of opportunities is your organization considering for incorporating AI into cybersecurity defensive operations?
(4/4)

- developed in-house
- Shark out the bad actor
- Communications to clients (Reporting)
- none yet
- -
- Financial market analysis Scrubbing internal data
- OWASP AI projects
- We will consider treat detection and analysis and incident response using AI
- Incorporation into Operation Technology and SCADA Monitoring for breaches and low level attack vectors.
- defining what AI is used for
- Attack and Pen tests
- Predictive and proactive modeling and deterrence
- cyber attacjs
- Automated pen-testing
- none

Slido Results: *Conducting AI-Enabled Cyber Defense 12 of 17)*

Conducting AI-enabled Cyber Defense (3/4)

0 3 2

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations?
(1/5)

- 1) data flow when AI sees our cyber data. any 3rd party flow is involved 2) what AI can remember from our sensitive cyber data, if the model is shared with others 3) how often AI needs to be certified if I am using AI for my cyber defense
- Ensure the AI is in compliance
- Clear definition of the AI cyber controls and reconnected solutions of how to implement. Also, strategies on how to build continual improvement into operational those controls.
- Human / AI interaction LLM safety/explainability
- Change in human behavior due to AI used elsewhere in the organization.
- I think it's important to build a SWOT for AI Profiles.
- Fraud detection

Slido Results: *Conducting AI-Enabled Cyber Defense 13 of 17)*

Conducting AI-enabled Cyber Defense (3/4)

0 3 2

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations?
(2/5)

- Really depends by specific feature or area. for this question. May want to re-state or re-write.
- I have no opinion
- Efforts to close the Skills gap between Level 1 - Level 3 by having faster insertion of high level expertise earlier in the Incident Response
- Cycle; Use AI to accelerate information sharing on real time threats with anonymized data until the speed of information sharing is real time
- Prioritize proticol standards, response tools and support for for generative Ai models that result in consumer facing

Slido Results: *Conducting AI-Enabled Cyber Defense 14 of 17)*

Conducting AI-enabled Cyber Defense (3/4)

0 3 2

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations?
(3/5)

- end products. Build security standards into Ai models standards for development - with governance tools for safety check.
- The option answers procedure process instructions
- None I can think of
- how to integrate CSF 2 and NIST AI RMF to the US AI Action Plan
- Treat AI as Artificial Persons with Cognitive Capability.
- insider threat
- not sure
- Need to know when AI is actually in a tool. When it is, how was it trained, etc? Related to shadow AI, but that also includes employees using AI-enabled tools.
- The guy just managed attach vectors. That's a big problem we have

Slido Results: *Conducting AI-Enabled Cyber Defense 15 of 17)*

Conducting AI-enabled Cyber Defense (3/4)

0 3 2

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations?
(4/5)

- people using anthropomorphic verbiage to promote ideas that are additional attack vectors.
- Not exactly sure, but I would add that we probably need to consider security of the AI components themselves.
- Not sure
- Agentic AI
- We still need to tidy up the classical code. Seeing it's still vulnerable.
- SBOM for AI
- the growth of AI, every year AI growth and become more complex
- Too many.
- Data leakage Agentic Misalignment (https://www.anthropic.com/research/agentic-misalignment)
- Secure AI training
- Our big challenge area that does not

Slido Results: *Conducting AI-Enabled Cyber Defense 16 of 17)*

Conducting AI-enabled Cyber Defense (3/4)

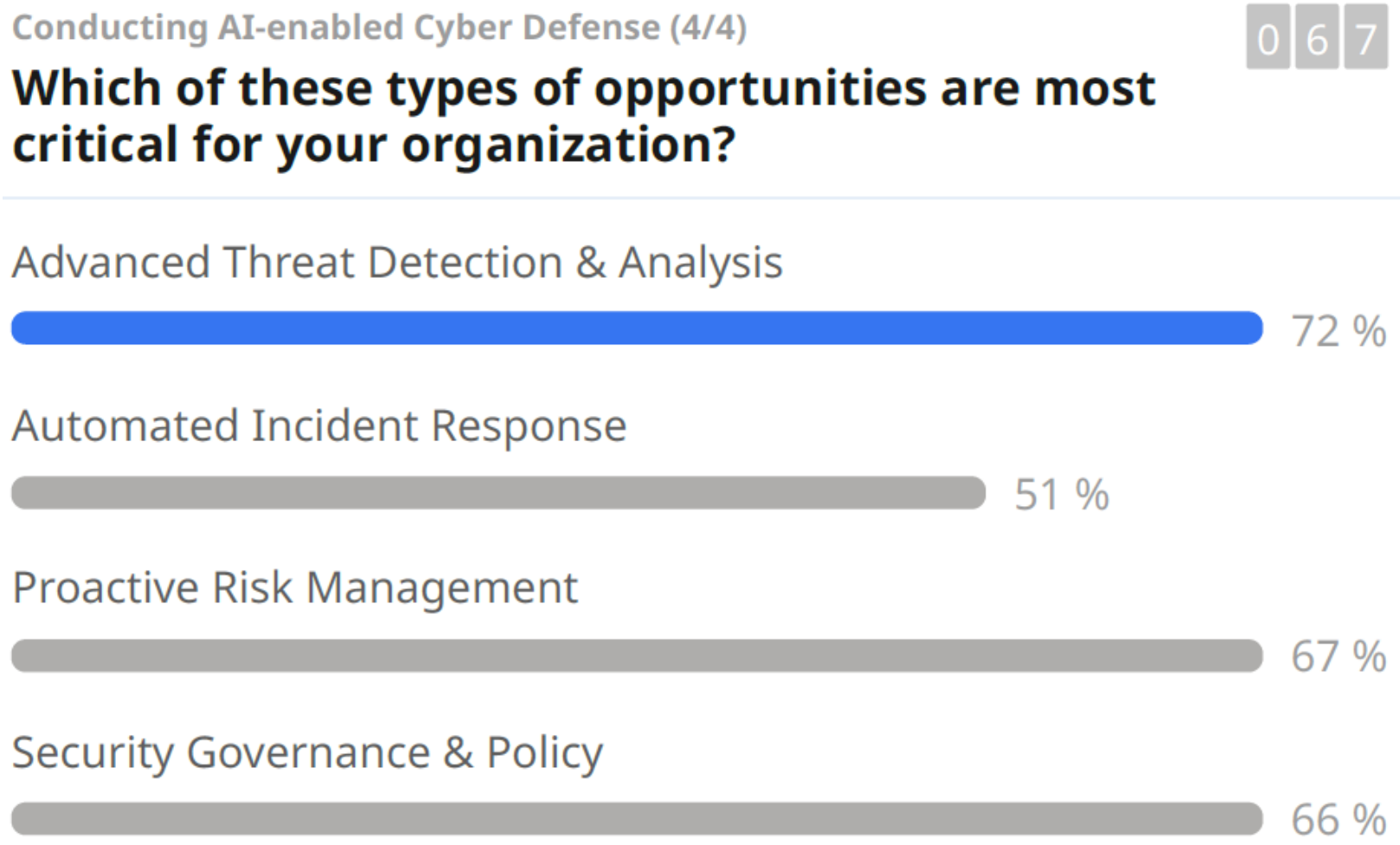
0 3 2

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations?
(5/5)

seem to be there is data labeling.

- See above
- not sure
- none

Slido Results: *Conducting AI-Enabled Cyber Defense 17 of 17)*



General Discussion Plan

- Walk through each CSF Function using Slido to foster discussion
- Questions we would like to address for each Function:
  - What are the:
    - Unique implications for the use of AI in cyber defense?
    - Most critical mitigations?
  - Based on the heatmaps:
    - Are the necessary Categories emphasized? Which Categories are over/under emphasized and why?
    - How well does the heatmap reflect current practices or other necessary outcomes?
  - Are there other important outcomes that are not represented?
  - Where do you need additional guidance, examples, or implementation resources to help your organization adopt AI-enabled technologies?
  - What resources are available to inform priorities (e.g., standards, mappings, tools)?

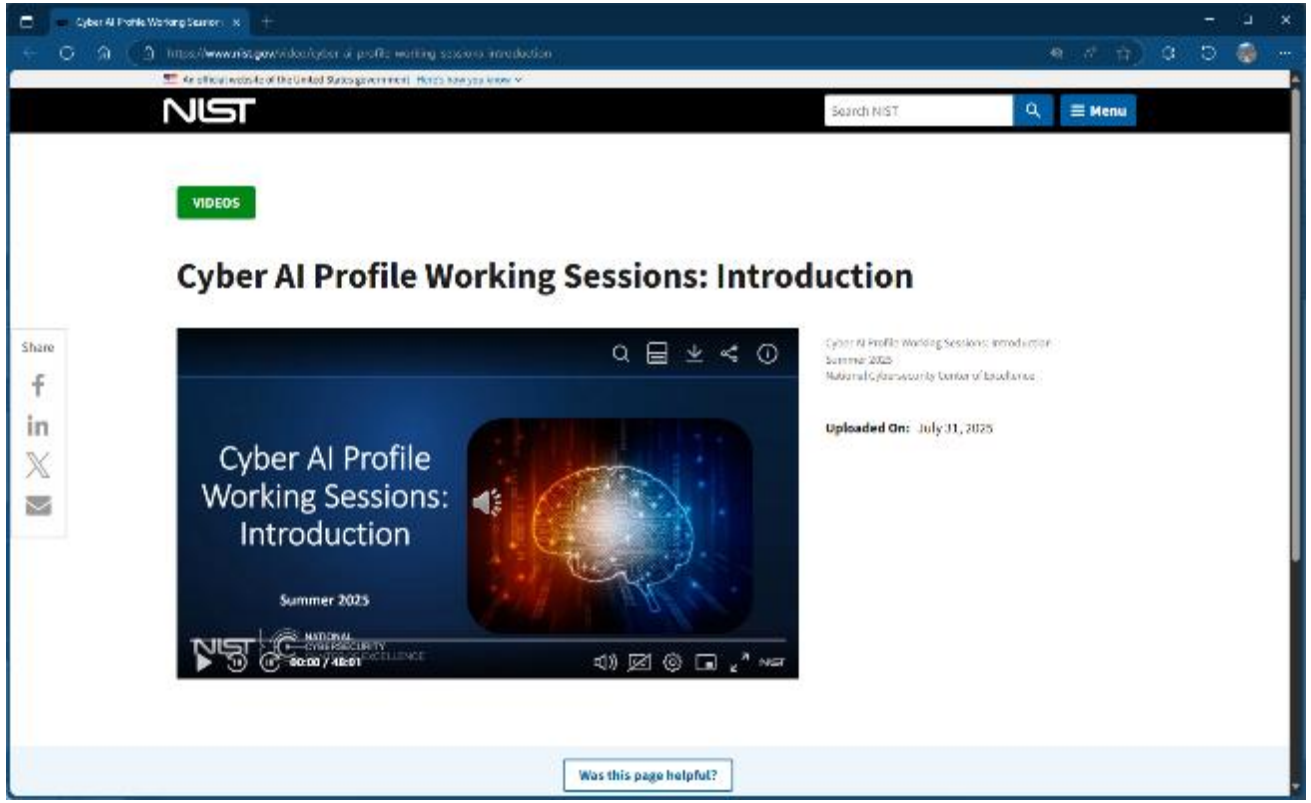
CSF 2.0 PDF

Cybersecurity Framework page

Refresher from Working Session Introduction

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Working Sessions Introduction Video



To help us maximize our working time during these sessions, we recorded an introduction video to provide background for anyone that is new to this process. The recording includes the following topics:

- Introduction to the NCCoE
- Background and Purpose for the Cyber AI Profile
- Overview of the NIST Cybersecurity Framework (CSF) 2.0
- Overview of Community Profiles
- Summary of Feedback in Early 2025
- Working Session Approach
- Resources

NIST CSF 2.0 Components

High-level hierarchy of cybersecurity outcomes that enable an organization to discuss and flexibly manage risk

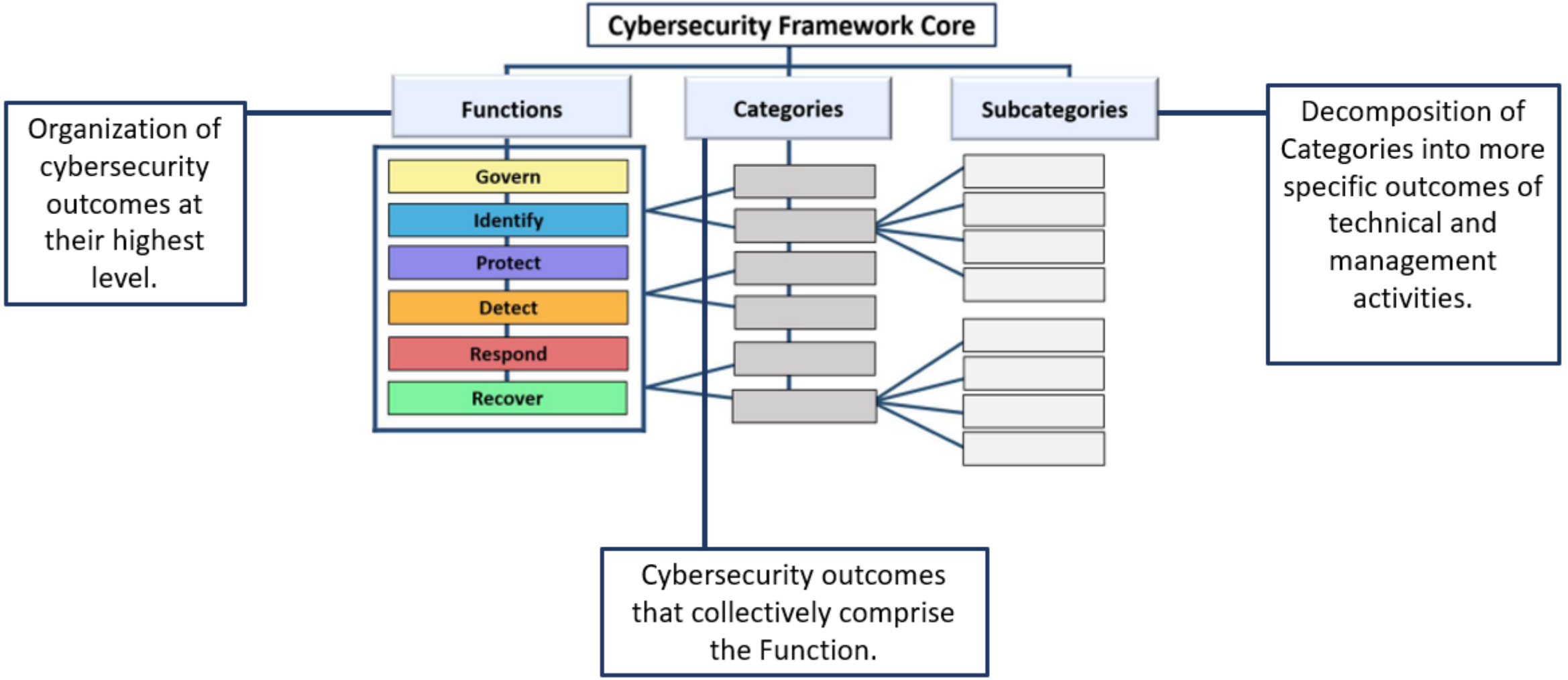
Help characterize the context and rigor of an organization's cybersecurity risk governance and management practices

Provide a way to understand, tailor, assess, prioritize, and communicate the Core's outcomes based on mission objectives, stakeholder expectations, threat landscape, and requirements

Each Component reinforces the connection between mission/business goals and cybersecurity outcomes.

Source (adapted): National Institute of Standards and Technology (2024) The NIST Cybersecurity Framework (CSF) 2.0. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) NIST CSWP 29. https://doi.org/10.6028/NIST.CSWP.29

NIST CSF 2.0 Core Structure



Source: National Institute of Standards and Technology (2024) The NIST Cybersecurity Framework (CSF) 2.0. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) NIST CSWP 29. https://doi.org/10.6028/NIST.CSWP.29

Notional Example Format

Assumption: The organization already has a cybersecurity program in place

Profile: Supplements the cybersecurity program by contemplating the unique cybersecurity risk management considerations that arise for each of the Cyber AI Profile Focus Areas

CSF Core	Securing AI System Components	Thwarting AI-enabled Cyber Attacks	Conducting AI-enabled Cyber Defense	Informative References / Mappings
CSF.XX-01: [Subcategory text]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[Pointers to related, laws, regulations, guidance, mappings, etc.]
CSF.XX-02: [Subcategory text]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[Pointers to related, laws, regulations, guidance, mappings, etc.]
CSF.XX-03: [Subcategory text]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[Pointers to related, laws, regulations, guidance, mappings, etc.]

Example Content - Extreme Fast Charging Profile (CSF 1.1)

CSF Core	Ecosystem-Wide	Electric Vehicles (EV)	eXtreme Fast Charging (XFC)/ Electric Vehicle Supply Equipment (EVSE)	Cloud/Third-Party Organizations	Utilities/Building Systems	Informative References / Mappings
GV-2: Cybersecurity roles and responsibilities are coordinated and aligned with internal roles and external partners.	Agreements with external organizations or partners are typically made in advance and documented in a service-level agreement (SLA), memorandum of understanding (MOU), or other forms of agreement. These agreements clearly define cybersecurity roles and responsibilities to properly define how their cybersecurity programs should function in a coordinated manner and allow for accountability for participant responsibilities.	Roles and responsibilities may include those involved in vehicle design, pre/post-sales support, software/firmware lifecycle activities, and supporting nominal vehicle operations such as charging, maintenance, and patching.	Roles and responsibilities may include those during EVSE installation design, construction, maintenance, updating, and operation. EVSE manufacturers can also consider defining roles to better support the needs of EV/XFC partners and customers, which may follow established OT or IT processes and methods for equipment, remote services, and capabilities.	Applicable, no additional Cloud/Third-Party-specific considerations.	Applicable, no additional Utility/Building Management System-specific considerations.	**Ecosystem:** [NIST-SP800-53r5] PM-1, PM-2, PM-29, PS-7, PS-9 **EV:** ISO/SAE 21434 RQ-07-04, RQ-07-06, WP-07-01 **SFC/EVSE:** ISA/IEC 62443-2- 1:D4E1 ORG 1.3 **Cloud/Third-Party:** [NIST-SP800-53r5] PM-1, PM-2, PM-29 **Utilities/Building Systems:** ISA/IEC 62443-2- 1:D4E1 ORG 1.3

CSF 2.0 Category Considerations: Conducting AI-enabled Cyber Defense

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Mapping AI Opportunities in Cyber Defense to NIST CSF 2.0

Opportunities	CSF 2.0 Function(s)	CSF 2.0 Categories
Advanced Threat Detection & Analysis	IDENTIFY PROTECT DETECT	ID.RA: Risk Assessment PR.AT: Awareness and Training DE.CM: Continuous Monitoring DE.AE: Adverse Event Analysis
Automated Incident Response	RESPOND	RS.MA: Incident Management RS.AN: Incident Analysis RS.CO: Incident Response Reporting and Communication RS.MI: Incident Mitigation
Proactive Risk Management	GOVERN IDENTIFY	GV.SC: Supply Chain Risk Management ID.AM: Asset Management ID.RA: Risk Assessment
Security Governance & Policy	GOVERN	GV.RM: Risk Management Strategy GV.PO: Policy

AI Cybersecurity Threats and Mitigations

- **Goal:** Build on growing body of AI cybersecurity mitigations to identify impactful CSF 2.0 Subcategories for the 3 Cyber AI Profile focus areas/priorities
- **Approach:** Constructed a "heatmap" based on various frameworks and best practices documents published by:
  - Research Organizations
  - Non-profit Organizations
  - Technology Companies
- **NOTE:** The heatmaps presented during these working sessions were developed as a tool for facilitating Cyber AI Profile development discussions and is not intended to be used for any other purpose.

Sources of Example Inputs

Concept Documents	Mapped Documents
• Cloud Security Alliance (CSA) • Center for Security and Emerging Technology (CSET) • Institute for Security + Technology (IST) • R Street	• Databricks • European Union Agency for Cybersecurity (ENISA) • Google • MITRE ATLAS™ • OWASP

Questions for discussion:

- What additional resources should be included?
- Are there critical mitigations that are missing from the current body of work?

Align Industry Mitigations to NIST CSF 2.0

Step 1:
Examine available publications

Step 2:
Assess whether the identified threats and mitigations are addressed by CSF 2.0

Step 3:
Align threat mitigations with CSF Subcategories and assess their coverage

Sources of Example Inputs

Concept Documents	Mapped Documents
• Cloud Security Alliance (CSA) • Center for Security and Emerging Technology (CSET) • Institute for Security + Technology (IST) • R Street	• Databricks • European Union Agency for Cybersecurity (enisa) • Google • MITRE ATLAS™ • OWASP


Security and Privacy Controls for Information Systems and Organizations

CSF Category Coverage		
Category	Count	Normalized
GV	160	0.5
ID	136	0.4
PR	315	1.0
DE	41	0.1
RS	13	0.0
RC	1	0.0

Legend	
Priority	Color
Low	
Medium	
High	

CSF Subcategory Coverage		
Subcategory	Count	Normalized
GV.OC	46	0.4
GV.RM	17	
GV.RR	20	
GV.PO		
GV.OV		
GV.SC		
ID.AM		
ID.RA		
ID.IM		
PR.AA		0.3
PR.AT		0.4
PR.DS	117	1.0
PR.PS	96	0.5
PR.IR	99	0.5
DE.CM	24	0.2
DE.AE	17	0.1
RS.MA	8	0.1
RS.AN	1	0.0
RS.CO	6	0.1
RS.MI	0	0.0
RC.RP	1	0.0
RC.CO	0	0.0

FOR DISCUSSION PURPOSES ONLY

Summary View

GOVERN	Heatmap
Organizational Context (GV.OC)	0.5
Risk Management Strategy (GV.RM)	0.5
Roles, Responsibilities, and Authorities (GV.RR)	0.1
Policy (GV.PO)	0.1
Oversight (GV.OV)	0.2
Cybersecurity Supply Chain Risk Management (GV.SC)	0.9

IDENTIFY	Heatmap
Asset Management (ID.AM)	0.7
Risk Assessment (ID.RA)	1.0
Improvement (ID.IM)	0.5

PROTECT	Heatmap
Identity Management, Authentication and Access Control (PR.AA)	0.5
Awareness and Training (PR.AT)	0.1
Data Security (PR.DS)	0.5
Platform Security (PR.PS)	0.6
Technology Infrastructure Resilience (PR.IR)	0.3

DETECT	Heatmap
Continuous Monitoring (DE.CM)	0.7
Adverse Event Analysis (DE.AE)	1.0

RESPOND	Heatmap
Incident Management (RS.MA)	0.3
Incident Analysis (RS.AN)	0.1
Incident Response Reporting and Communication (RS.CO)	0.4
Incident Mitigation (RS.MI)	0.1

RECOVER	Heatmap
Incident Recovery Plan Execution (RC.RP)	0.3
Incident Recovery Communications (RC.CO)	0.1

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of emphasis/potential priority):

Opportunities & Priorities in the CSF Core

Discussion Flow:

- Walk through each CSF Function
- Discussion questions:
  - How can AI support organizations with achieving the outcomes in the Function?
  - Does the heatmap emphasize the right degree of priority for each Category?
  - Are there other important cybersecurity activities or outcomes for AI-enabled cyber defensive opportunities that belong in this Function but are not represented by the Categories?
  - What resources are available to inform priorities for this Function (e.g., standards, mappings, tools)?

CSF 2.0 PDF

Cybersecurity Framework page

Govern: Opportunities

Category	Description
Organizational Context (GV.OC)	The circumstances — mission, stakeholder expectations, dependencies, and legal, regulatory, and contractual requirements — surrounding the organization's cybersecurity risk management decisions are understood
Risk Management Strategy (GV.RM)	The organization's priorities, constraints, risk tolerance and appetite statements, and assumptions are established, communicated, and used to support operational risk decisions
Roles, Responsibilities, and Authorities (GV.RR)	Cybersecurity roles, responsibilities, and authorities to foster accountability, performance assessment, and continuous improvement are established and communicated
Policy (GV.PO)	Organizational cybersecurity policy is established, communicated, and enforced
Oversight (GV.OV)	Results of organization-wide cybersecurity risk management activities and performance are used to inform, improve, and adjust the risk management strategy
Cybersecurity Supply Chain Risk Management (GV.SC)	Cyber supply chain risk management processes are identified, established, managed, monitored, and improved by organizational stakeholders

Join at
slido.com
#CyberAI_WS

Slido Results: *Govern: Opportunities (1 of 2)*

Slido Results: *Govern: Opportunities (2 of 2)*

GOVERN: In which of these Categories do you see opportunities for integrating AI capabilities to support cybersecurity defenses?
(2/2)

0 6 7

Cybersecurity Supply Chain Risk Management (GV.SC)

67 %

Govern: Priorities

Slido.com
#CyberAI_WS

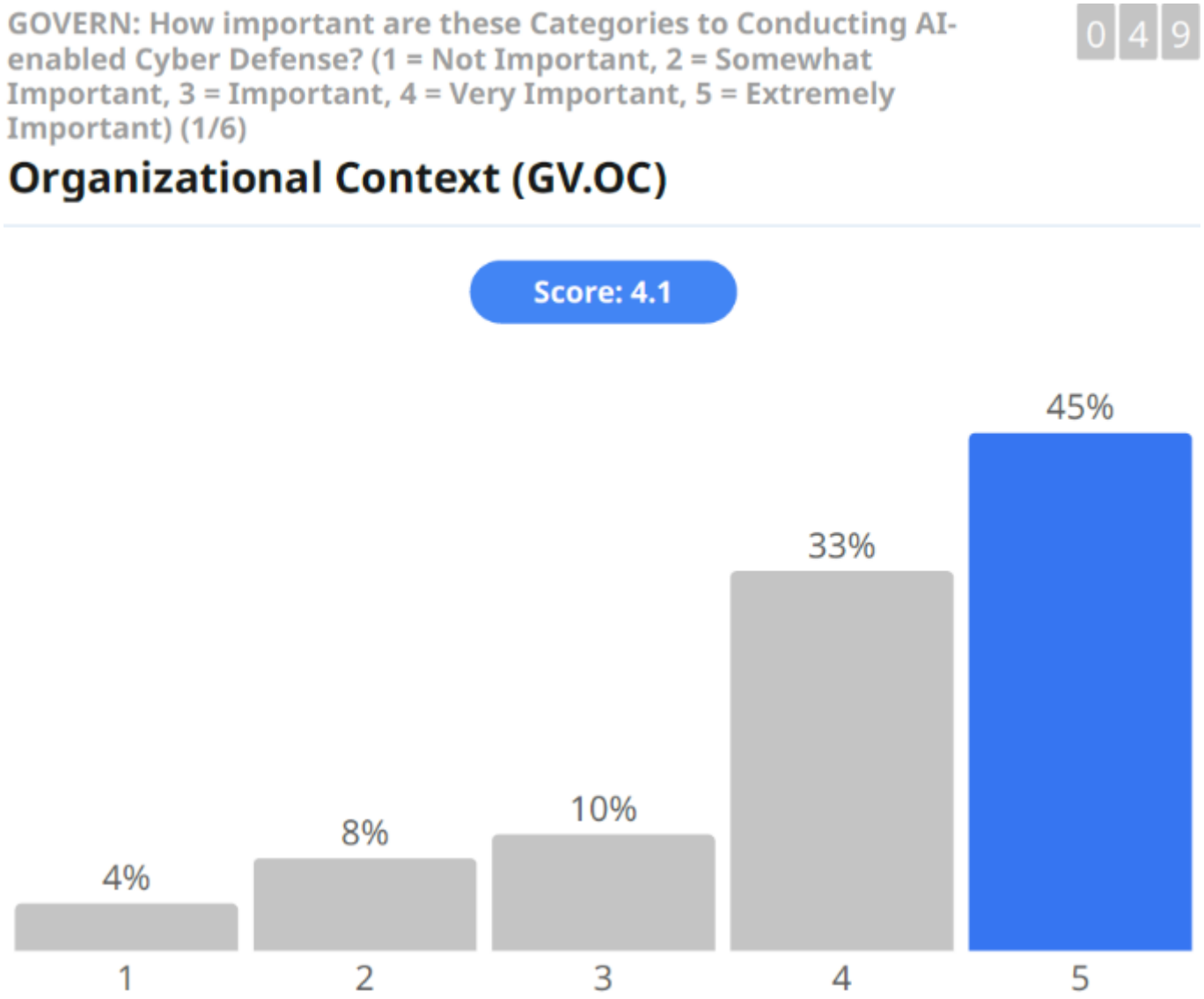
NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Category	Description	Heatmap
Organizational Context (GV.OC)	The circumstances — mission, stakeholder expectations, dependencies, and legal, regulatory, and contractual requirements — surrounding the organization's cybersecurity risk management decisions are understood	0.5
Risk Management Strategy (GV.RM)	The organization's priorities, constraints, risk tolerance and appetite statements, and assumptions are established, communicated, and used to support operational risk decisions	0.5
Roles, Responsibilities, and Authorities (GV.RR)	Cybersecurity roles, responsibilities, and authorities to foster accountability, performance assessment, and continuous improvement are established and communicated	0.1
Policy (GV.PO)	Organizational cybersecurity policy is established, communicated, and enforced	0.1
Oversight (GV.OV)	Results of organization-wide cybersecurity risk management activities and performance are used to inform, improve, and adjust the risk management strategy	0.2
Cybersecurity Supply Chain Risk Management (GV.SC)	Cyber supply chain risk management processes are identified, established, managed, monitored, and improved by organizational stakeholders	0.9

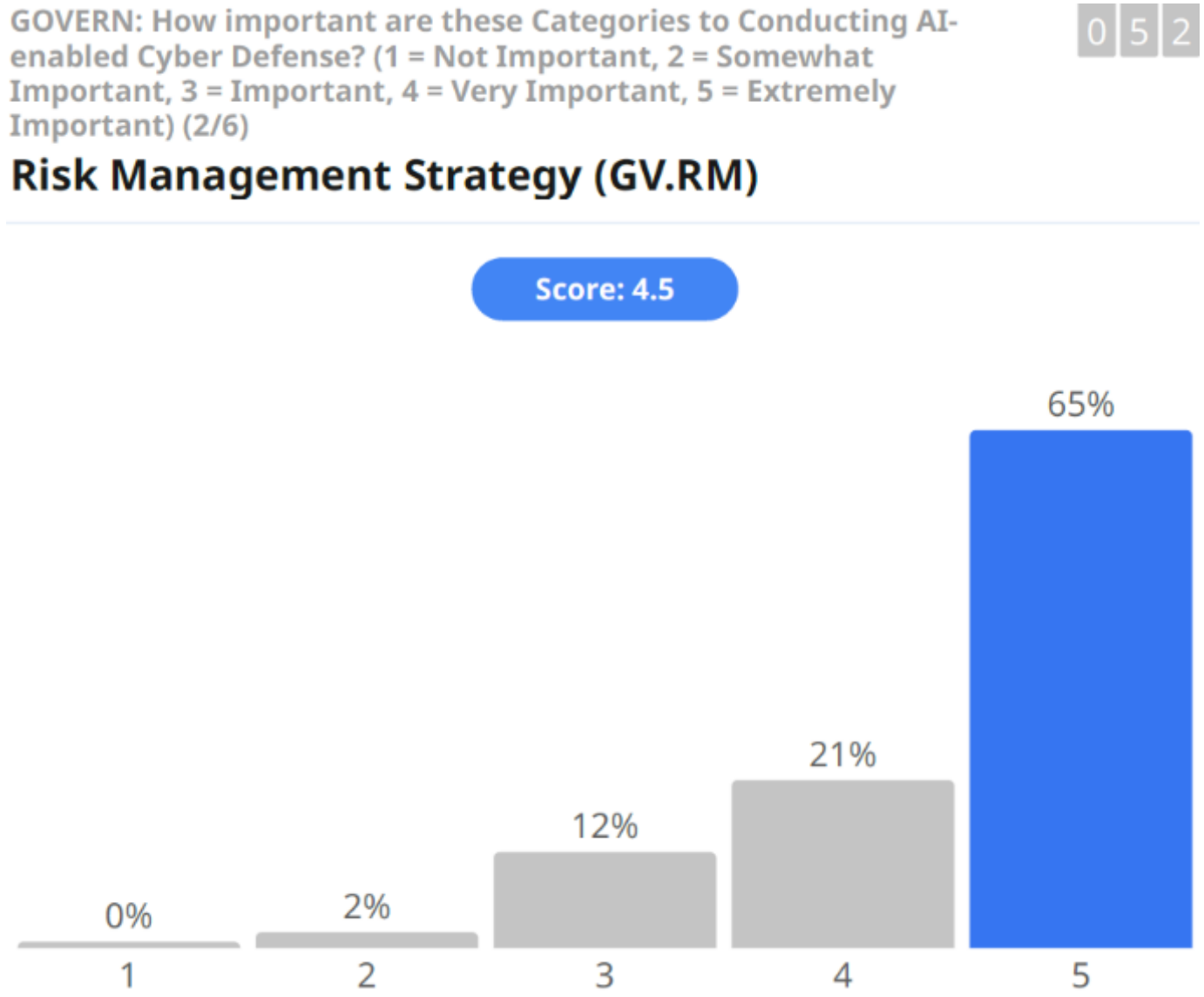
FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of emphasis/potential priority):

Slido Results: *Govern: Priorities (1 of 6)*



Slido Results: *Govern: Priorities (2 of 6)*



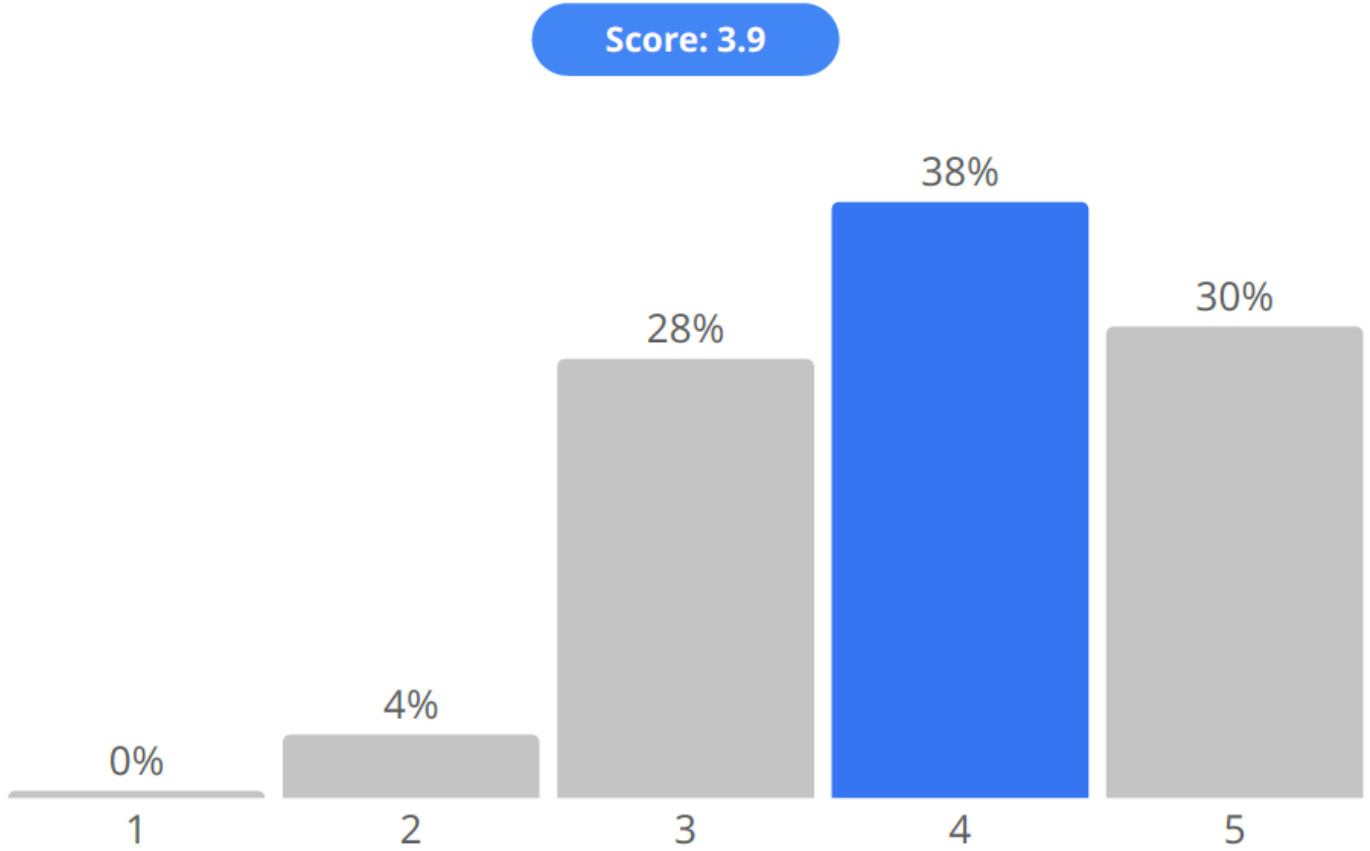
Slido Results: *Govern: Priorities (3 of 6)*

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

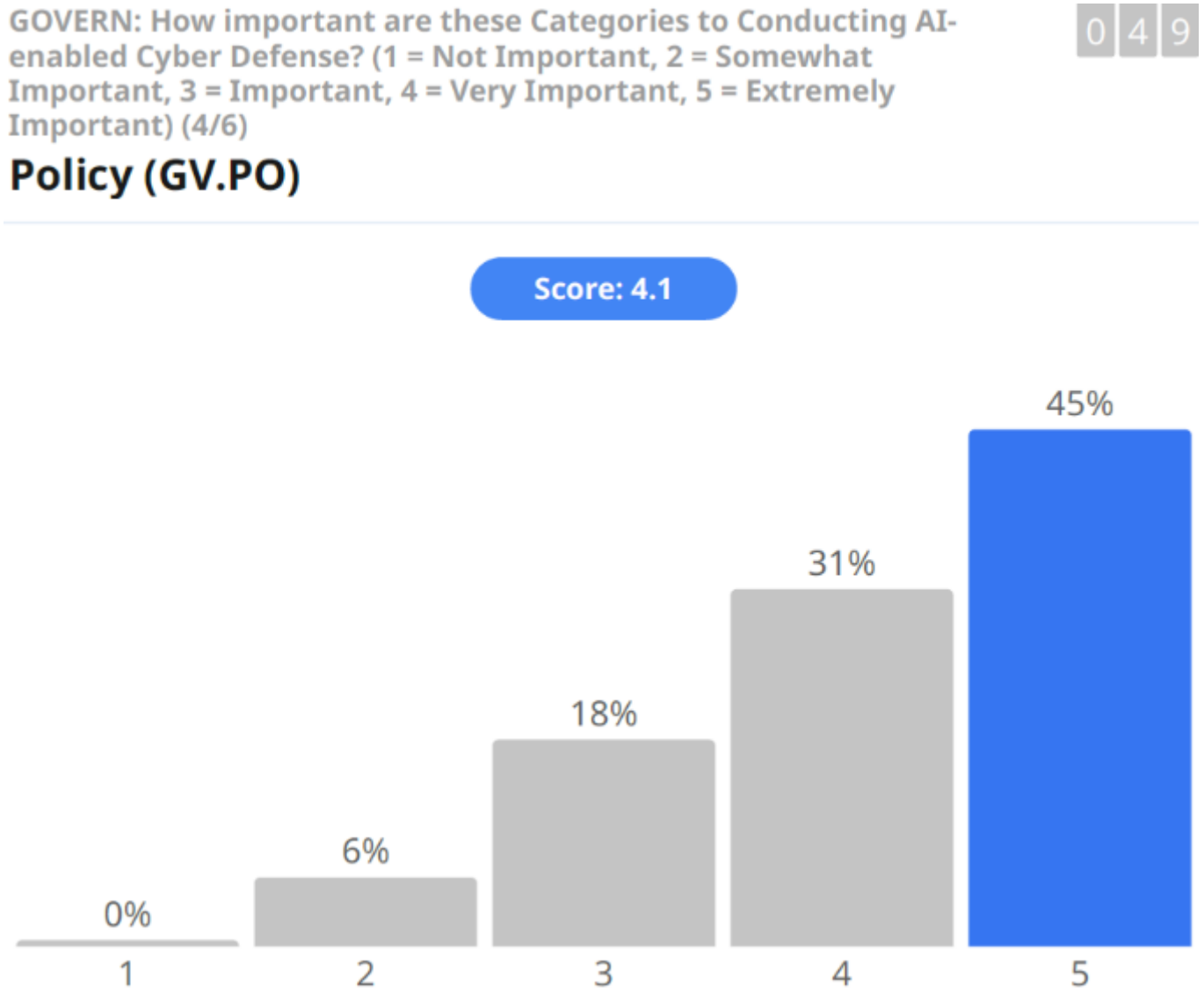
GOVERN: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (3/6)

0 5 0

Roles, Responsibilities, and Authorities (GV.RR)



Slido Results: *Govern: Priorities (4 of 6)*



Slido Results: *Govern: Priorities (5 of 6)*

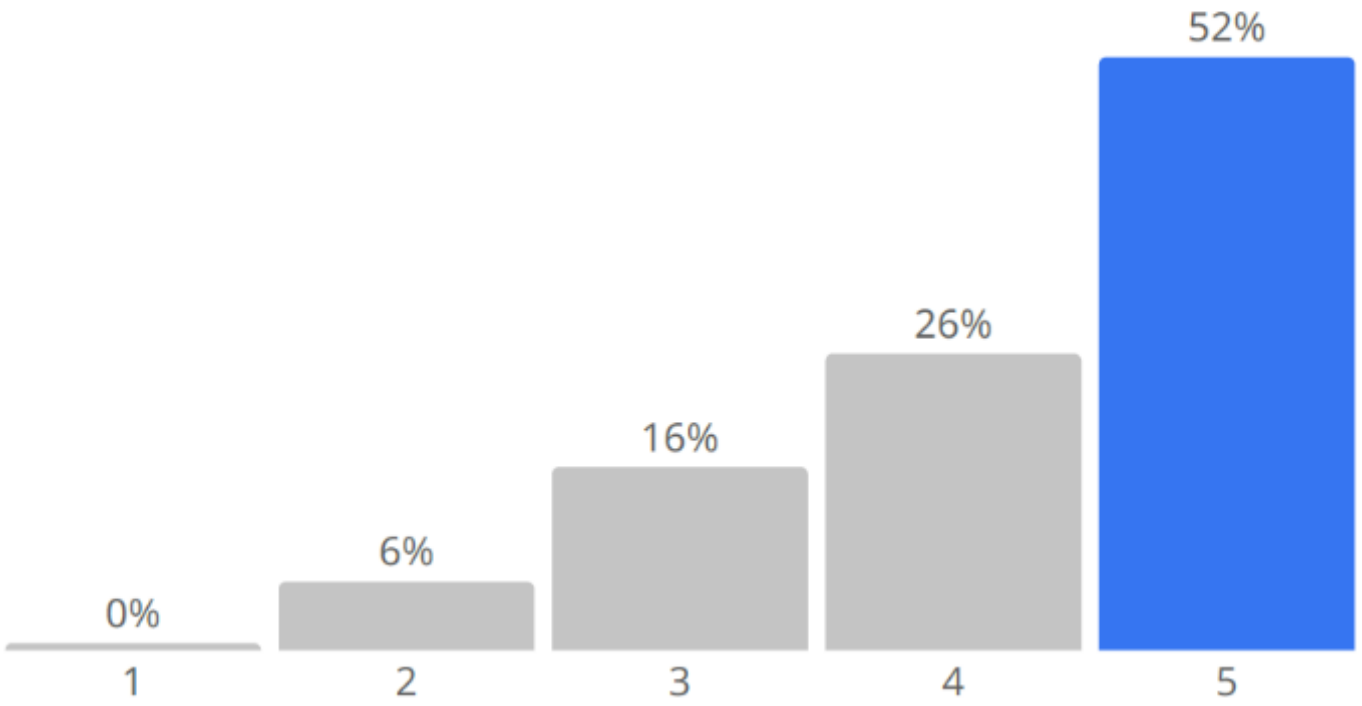
NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

GOVERN: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (5/6)

0 5 0

Oversight (GV.OV)

Score: 4.2



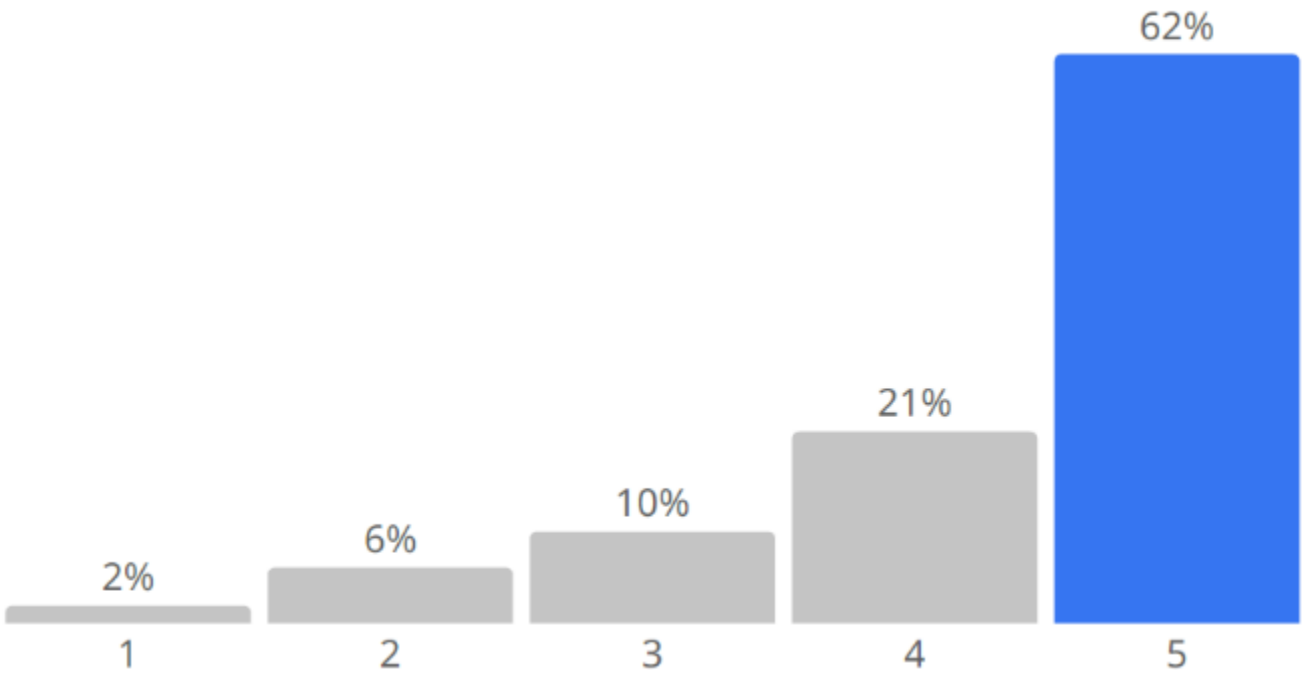
Slido Results: *Govern: Priorities (6 of 6)*

GOVERN: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (6/6)

0 5 2

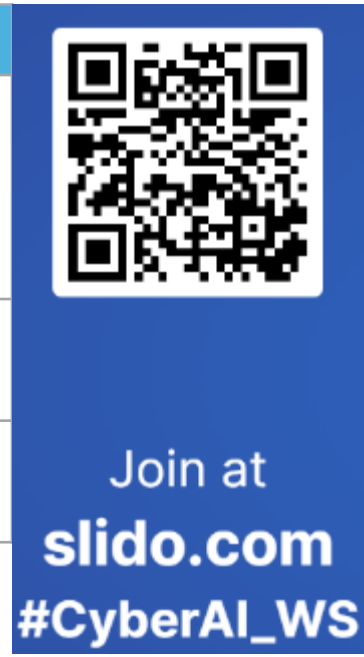
Cybersecurity Supply Chain Risk Management (GV.SC)

Score: 4.3

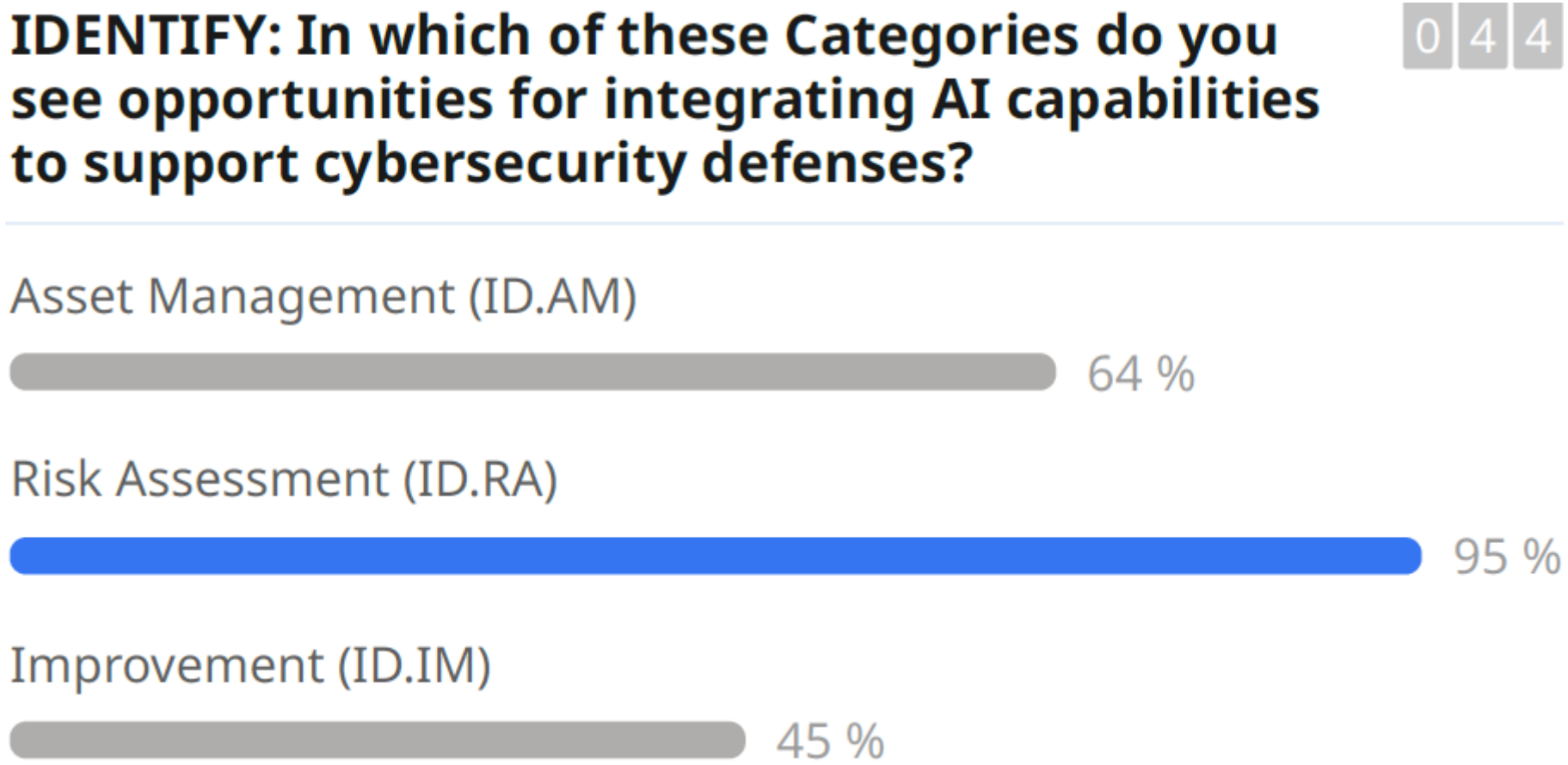


Identify: Opportunities

Category	Description
Asset Management (ID.AM)	Assets (e.g., data, hardware, software, systems, facilities, services, people) that enable the organization to achieve business purposes are identified and managed consistent with their relative importance to organizational objectives and the organization's risk strategy.
Risk Assessment (ID.RA)	The cybersecurity risk to the organization, assets, and individuals is understood by the organization.
Improvement (ID.IM)	Improvements to organizational cybersecurity risk management processes, procedures and activities are identified across all CSF Functions.



Slido Results: *Identify: Opportunities*



Identify: Priorities

Slido.com
#CyberAI_WS

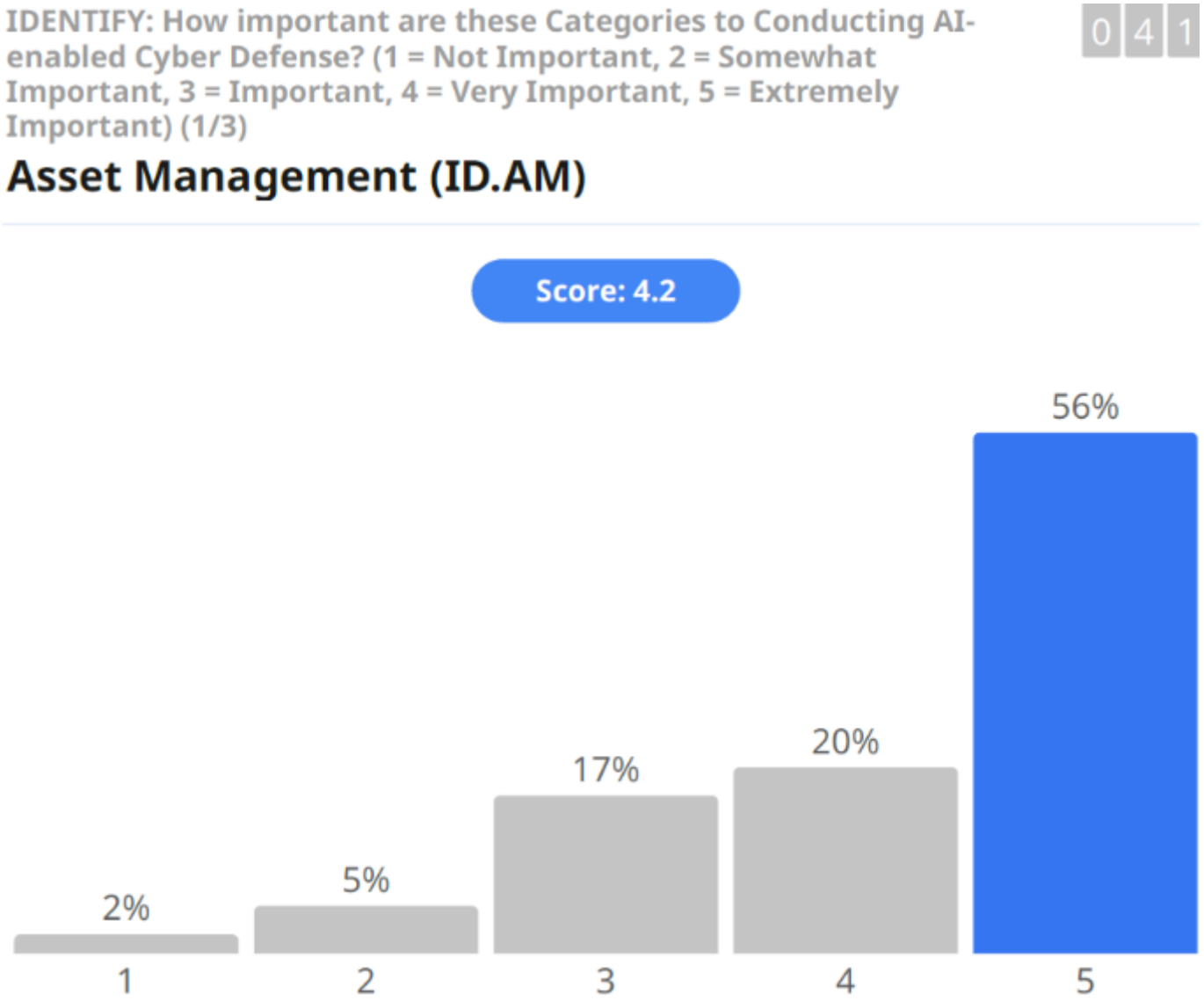
NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Category	Description	Heatmap
Asset Management (ID.AM)	Assets (e.g., data, hardware, software, systems, facilities, services, people) that enable the organization to achieve business purposes are identified and managed consistent with their relative importance to organizational objectives and the organization's risk strategy.	0.7
Risk Assessment (ID.RA)	The cybersecurity risk to the organization, assets, and individuals is understood by the organization.	1.0
Improvement (ID.IM)	Improvements to organizational cybersecurity risk management processes, procedures and activities are identified across all CSF Functions.	0.5

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of emphasis/potential priority):

Slido Results: *Identify: Priorities (1 of 3)*



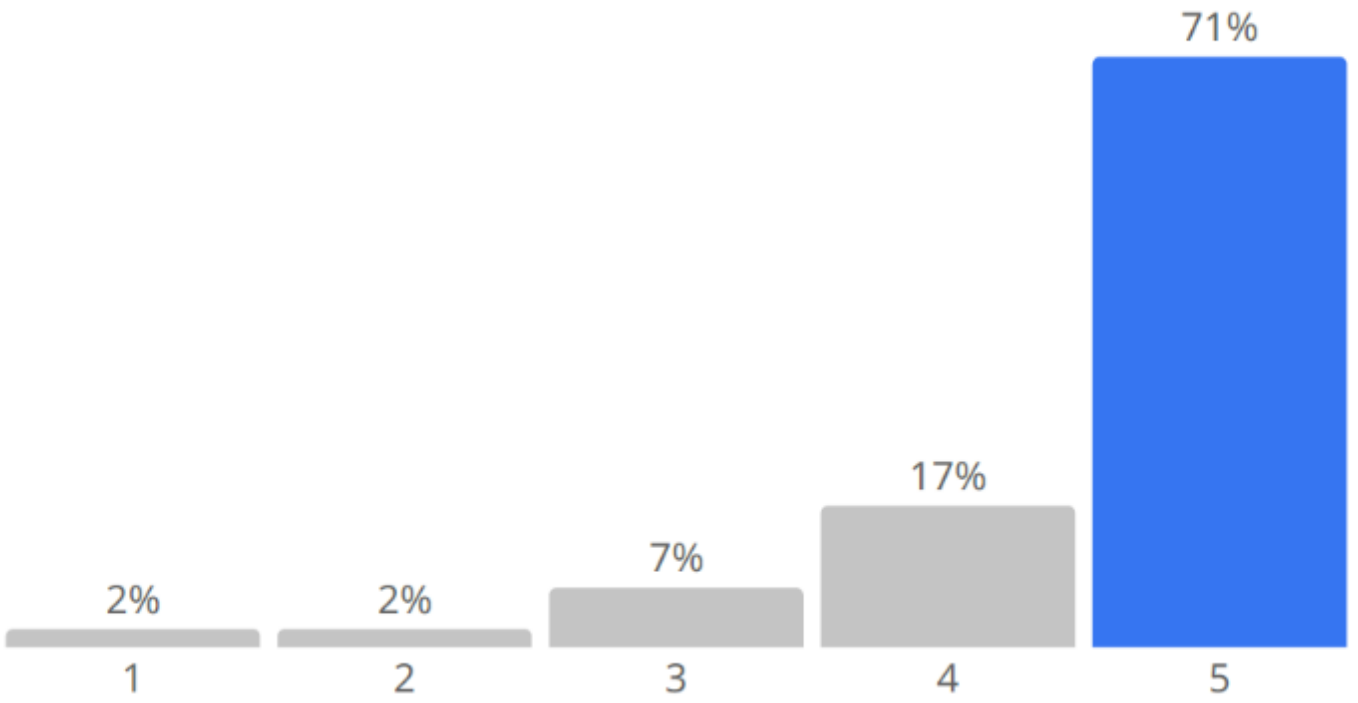
Slido Results: *Identify: Priorities (2 of 3)*

IDENTIFY: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (2/3)

0 4 1

Risk Assessment (ID.RA)

Score: 4.5

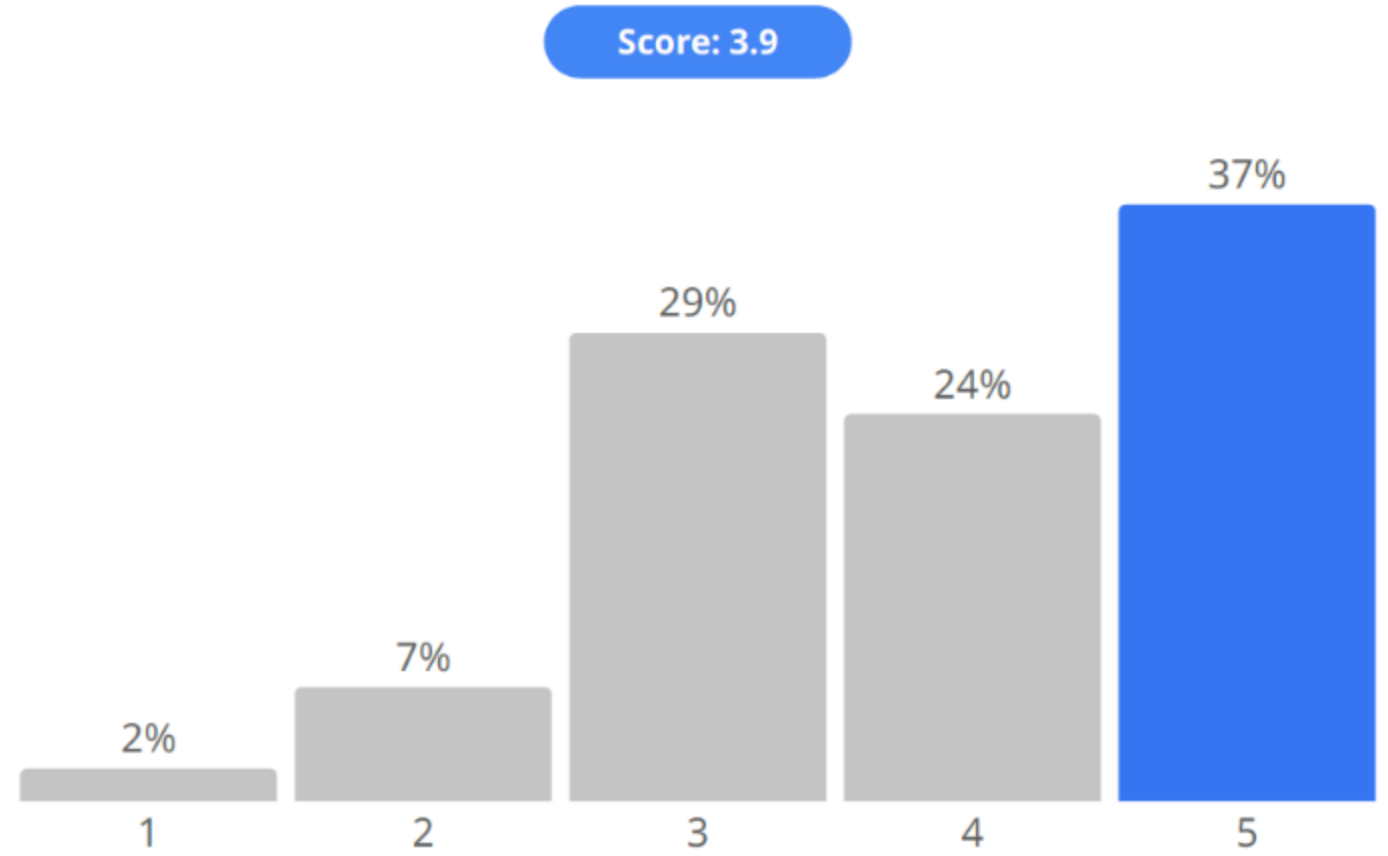


Slido Results: *Identify: Priorities (3 of 3)*

IDENTIFY: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (3/3)

0 4 1

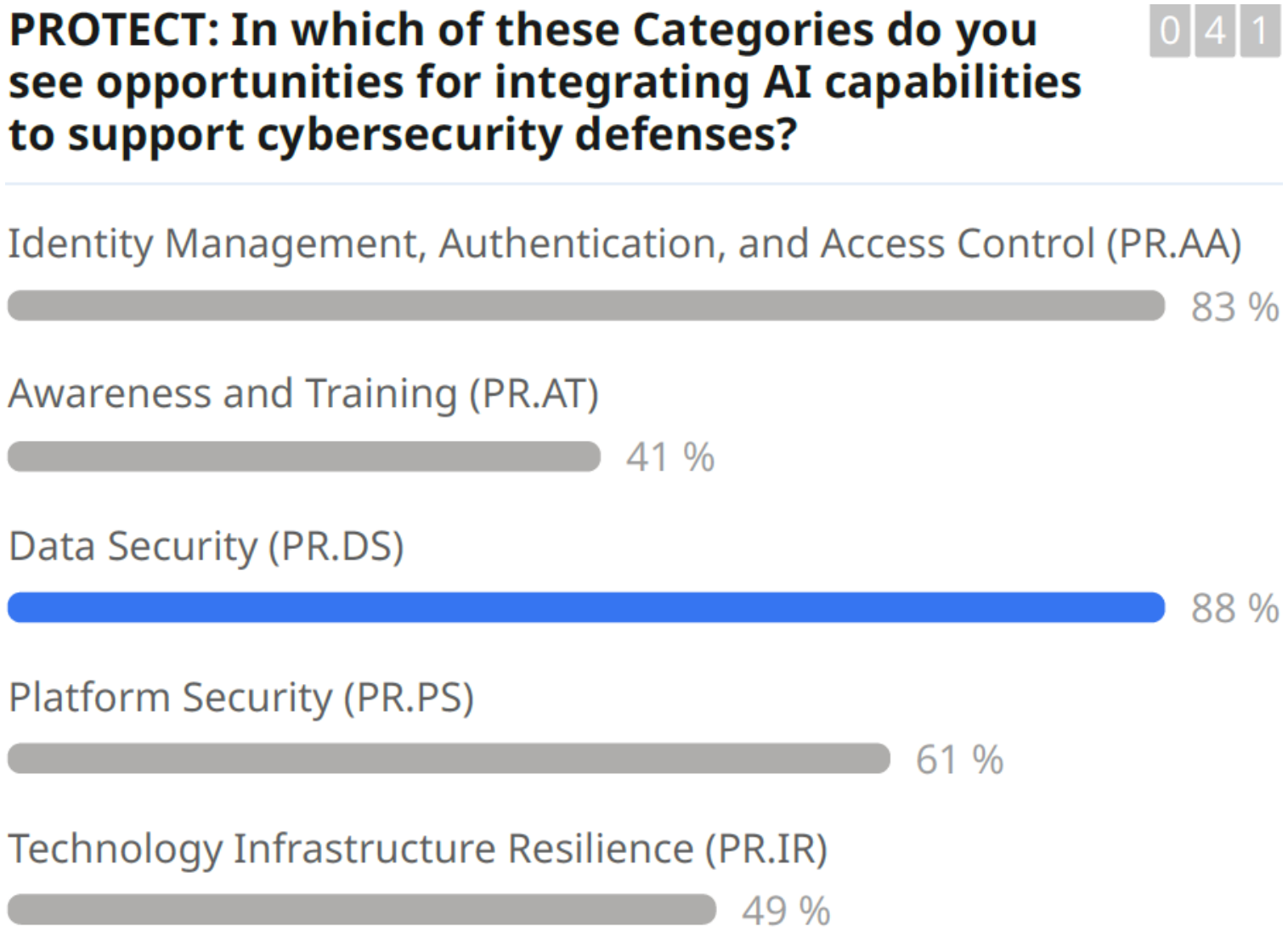
Improvement (ID.IM)



Protect: Opportunities

Category	Description
Identity Management, Authentication and Access Control (PR.AA)	Access to physical and logical assets is limited to authorized users, services, and hardware and managed commensurate with the assessed risk of unauthorized access.
Awareness and Training (PR.AT)	The organization's personnel are provided with cybersecurity awareness and training so that they can perform their cybersecurity-related tasks.
Data Security (PR.DS)	Data are managed consistent with the organization's risk strategy to protect the confidentiality, integrity, and availability of information.
Platform Security (PR.PS)	The hardware, software (e.g., firmware, operating systems, applications), and services of physical and virtual platforms are managed consistent with the organization's risk strategy to protect their confidentiality, integrity, and availability.
Technology Infrastructure Resilience (PR.IR)	Security architectures are managed with the organization's risk strategy to protect asset confidentiality, integrity, and availability, and organizational resilience.

Slido Results: *Protect: Opportunities*



Protect: Priorities

Slido.com
#CyberAI_WS

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Category	Description	Heatmap
Identity Management, Authentication and Access Control (PR.AA)	Access to physical and logical assets is limited to authorized users, services, and hardware and managed commensurate with the assessed risk of unauthorized access.	0.5
Awareness and Training (PR.AT)	The organization's personnel are provided with cybersecurity awareness and training so that they can perform their cybersecurity-related tasks.	0.1
Data Security (PR.DS)	Data are managed consistent with the organization's risk strategy to protect the confidentiality, integrity, and availability of information.	0.5
Platform Security (PR.PS)	The hardware, software (e.g., firmware, operating systems, applications), and services of physical and virtual platforms are managed consistent with the organization's risk strategy to protect their confidentiality, integrity, and availability.	0.6
Technology Infrastructure Resilience (PR.IR)	Security architectures are managed with the organization's risk strategy to protect asset confidentiality, integrity, and availability, and organizational resilience.	0.3

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of emphasis/potential priority):

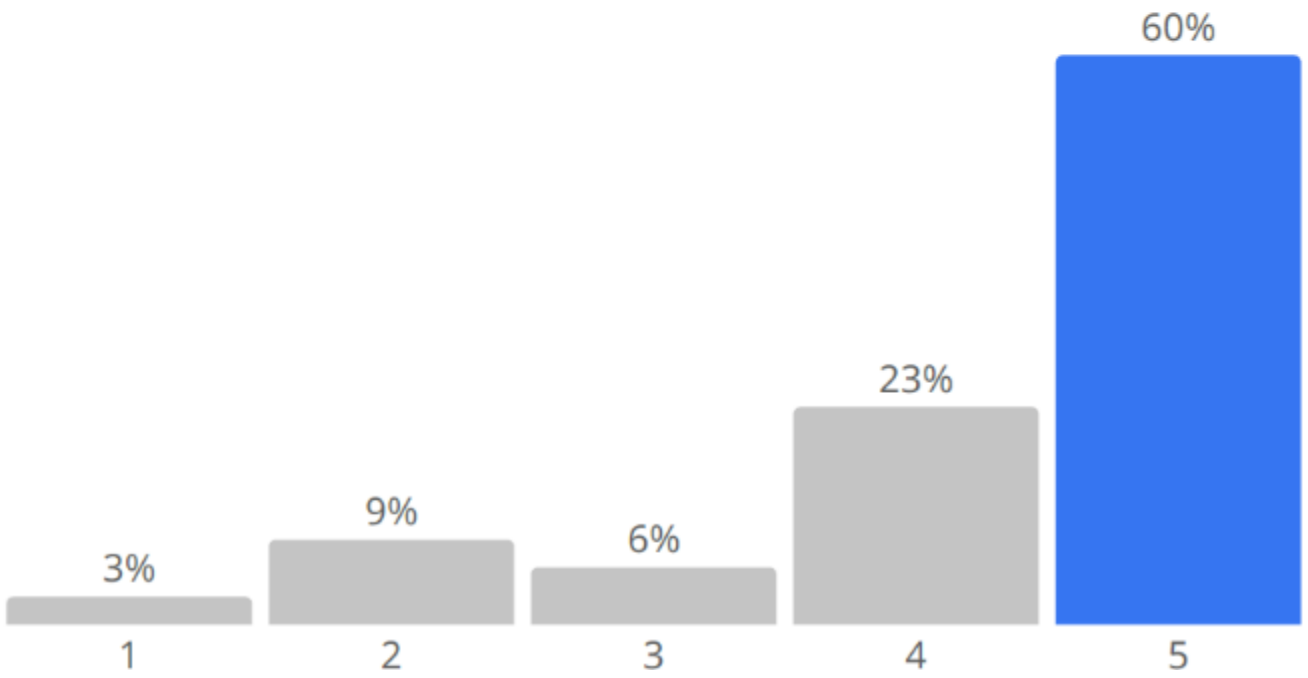
Slido Results: *Protect: Priorities (1 of 5)*

PROTECT: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (1/5)

0 3 5

Identity Management, Authentication, and Access Control (PR.AA)

Score: 4.3

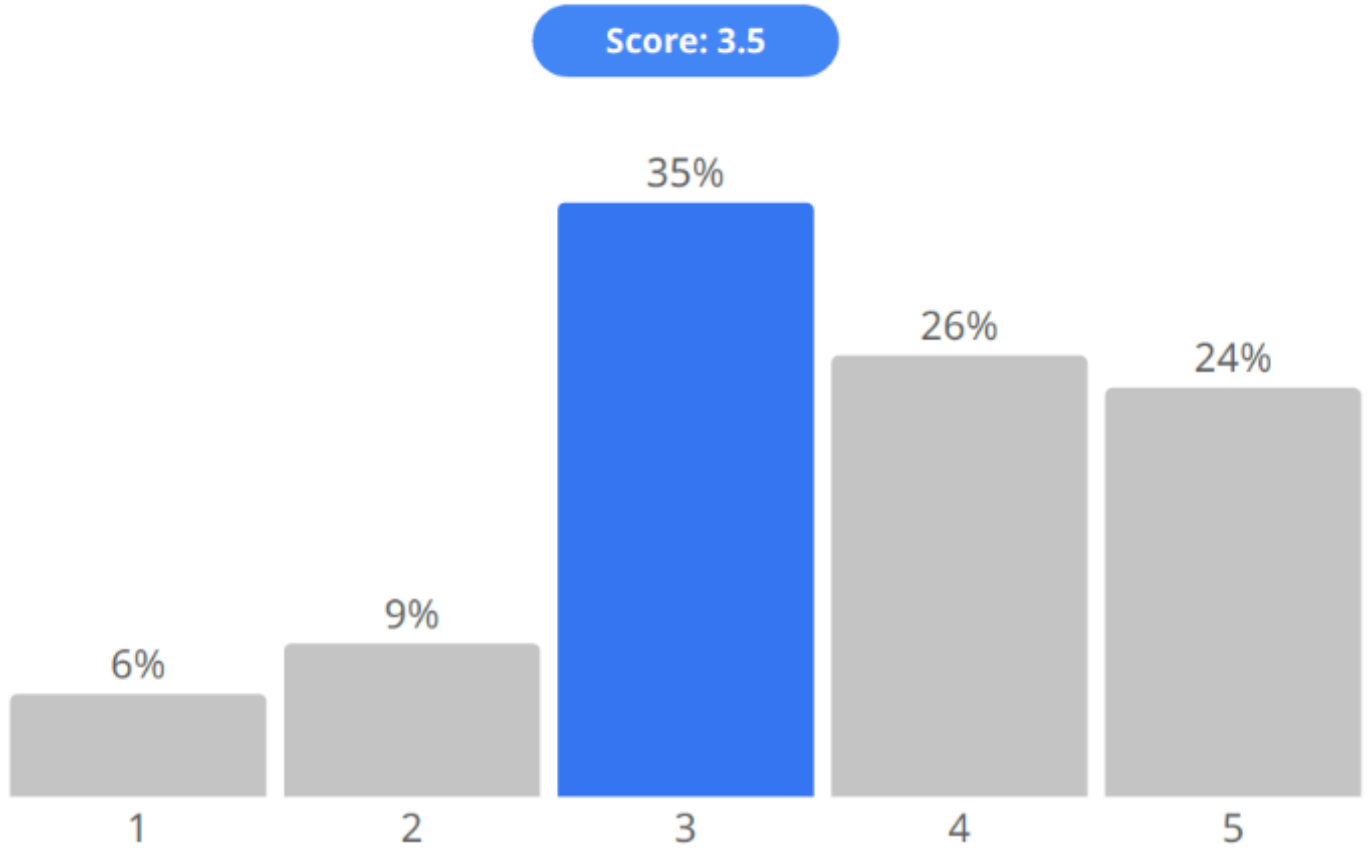


Slido Results: *Protect: Priorities (2 of 5)*

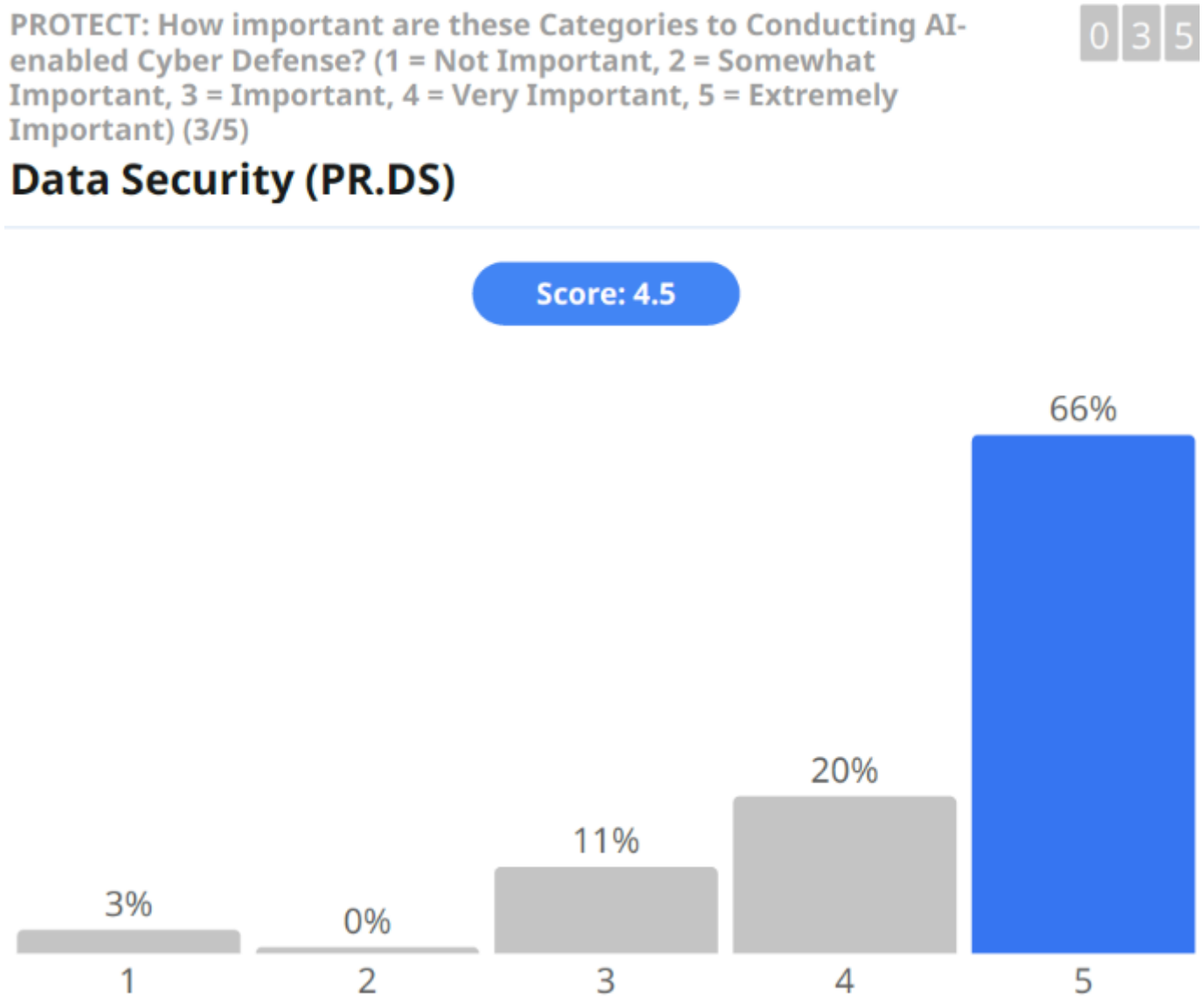
PROTECT: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (2/5)

0 3 4

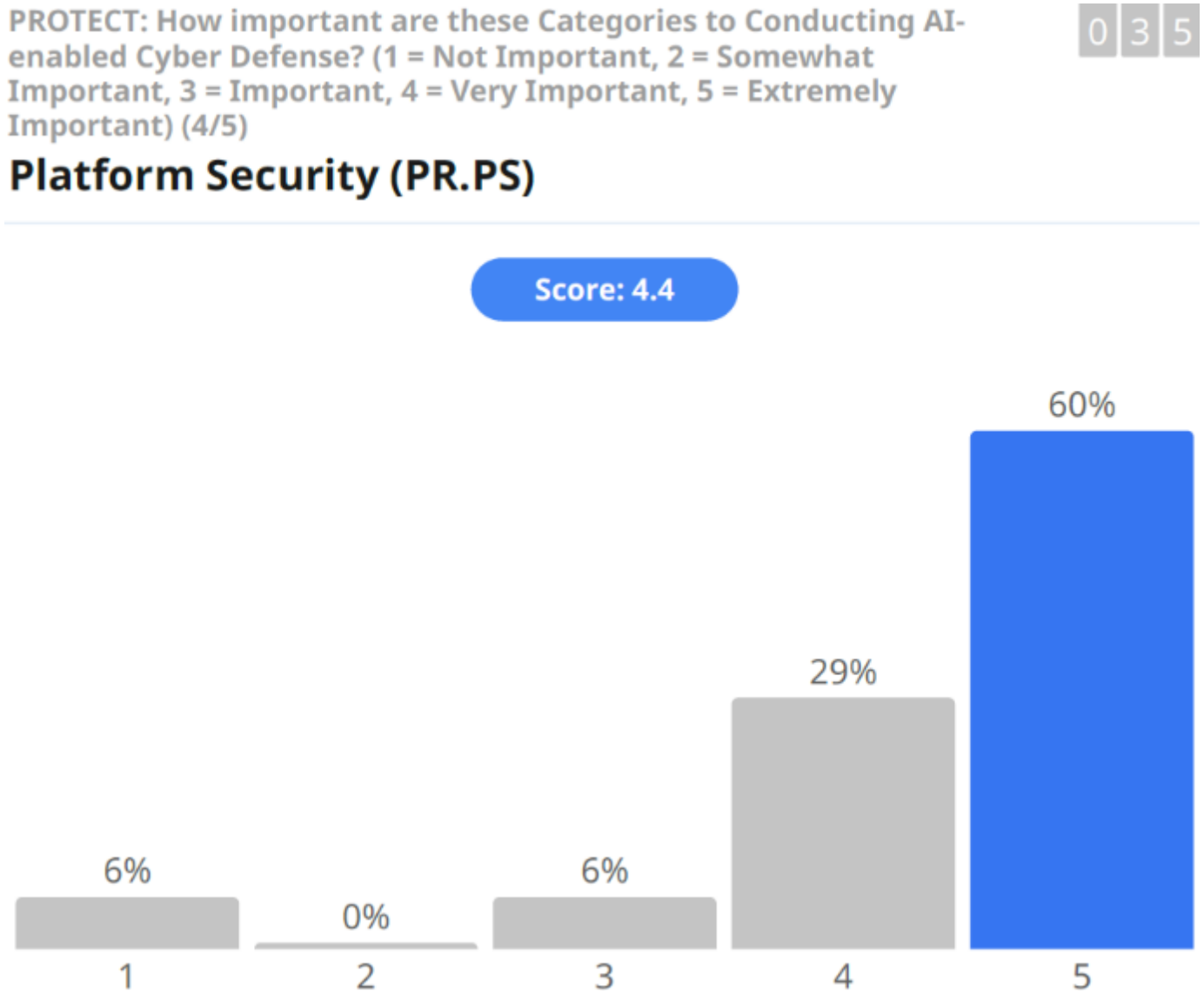
Awareness and Training (PR.AT)



Slido Results: *Protect: Priorities (3 of 5)*



Slido Results: *Protect: Priorities (4 of 5)*



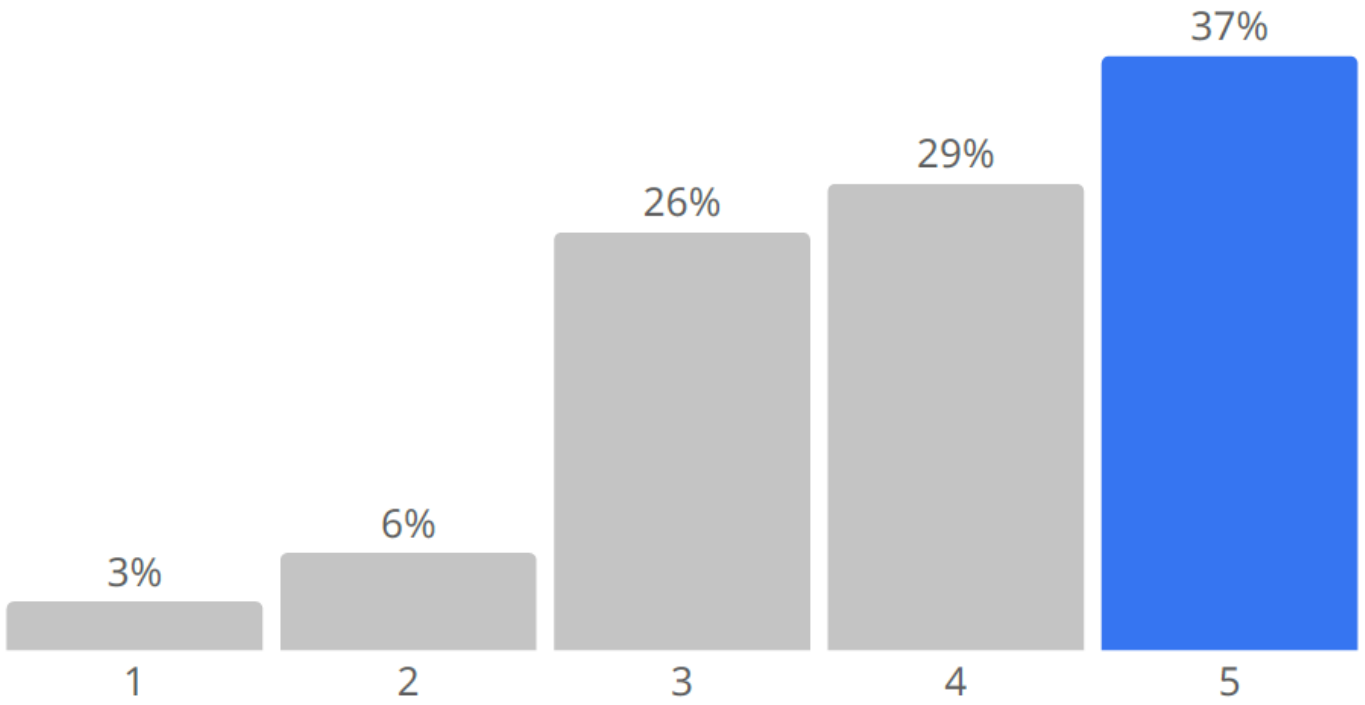
Slido Results: *Protect: Priorities (5 of 5)*

PROTECT: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (5/5)

0 3 5

Technology Infrastructure Resilience (PR.IR)

Score: 3.9



Detect: Opportunities

Category	Description
Continuous Monitoring (DE.CM)	Assets are monitored to find anomalies, indicators of compromise, and other potentially adverse events.
Adverse Event Analysis (DE.AE)	Anomalies, indicators of compromise, and other potentially adverse events are analyzed to characterize the events and detect cybersecurity incidents.

Detect: Priorities

Slido.com
#CyberAI_WS

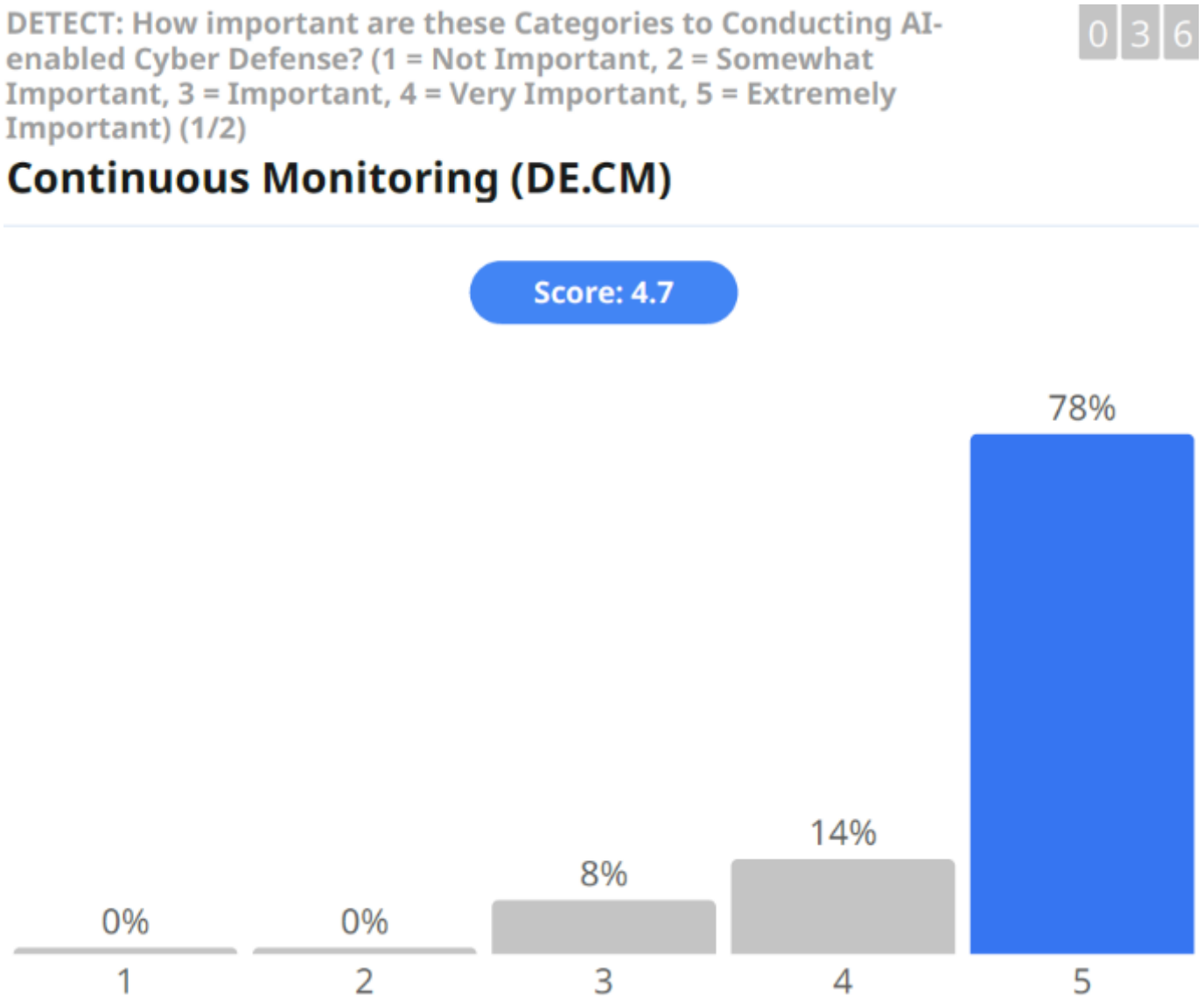
NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Category	Description	Heatmap
Continuous Monitoring (DE.CM)	Assets are monitored to find anomalies, indicators of compromise, and other potentially adverse events.	0.7
Adverse Event Analysis (DE.AE)	Anomalies, indicators of compromise, and other potentially adverse events are analyzed to characterize the events and detect cybersecurity incidents.	1.0

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of emphasis/potential priority):

Slido Results: *Detect: Priorities (1 of 2)*



Slido Results: *Detect: Priorities (2 of 2)*

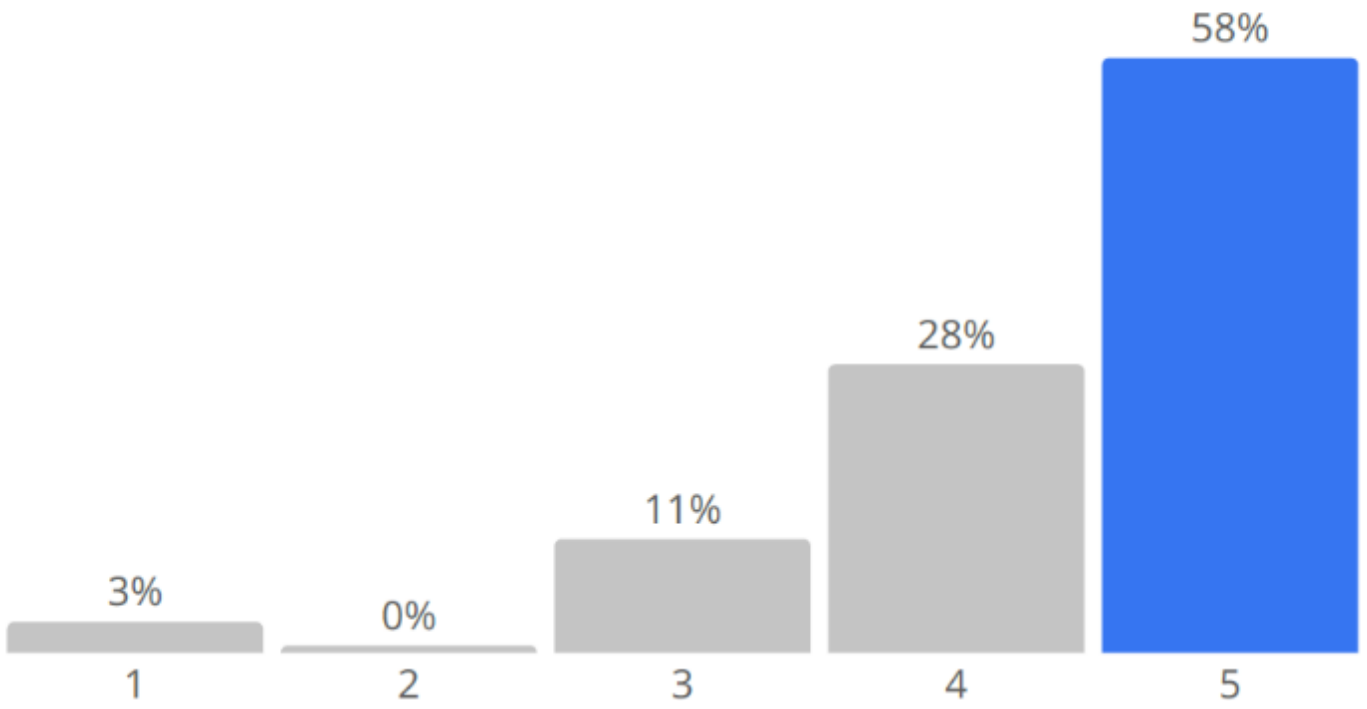
NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

DETECT: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (2/2)

0 3 6

Adverse Event Analysis (DE.AE)

Score: 4.4



Respond: Opportunities

Category	Description
Incident Management (RS.MA)	Responses to detected cybersecurity incidents are managed.
Incident Analysis (RS.AN)	Investigations are conducted to ensure effective response and support forensics and recovery activities.
Incident Response Reporting and Communication (RS.CO)	Response activities are coordinated with internal and external stakeholders as required by laws, regulations, or policies.
Incident Mitigation (RS.MI)	Activities are performed to prevent expansion of an event and mitigate its effects.

Respond: Priorities

Slido.com
#CyberAI_WS

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Category	Description	Heatmap
Incident Management (RS.MA)	Responses to detected cybersecurity incidents are managed.	0.3
Incident Analysis (RS.AN)	Investigations are conducted to ensure effective response and support forensics and recovery activities.	0.1
Incident Response Reporting and Communication (RS.CO)	Response activities are coordinated with internal and external stakeholders as required by laws, regulations, or policies.	0.4
Incident Mitigation (RS.MI)	Activities are performed to prevent expansion of an event and mitigate its effects.	0.1

FOR DISCUSSION PURPOSES ONLY

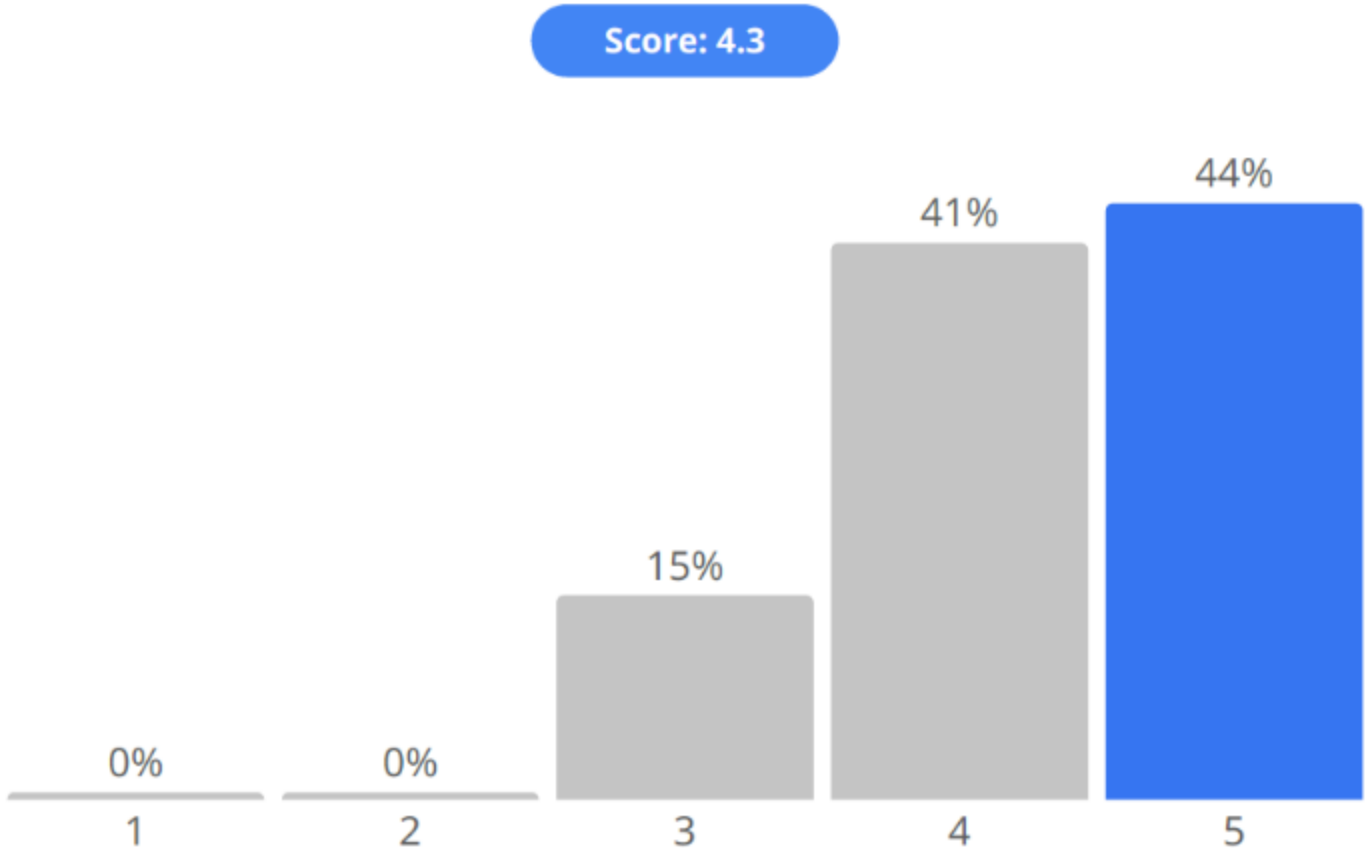
Heatmap Legend 0-1 (degree of emphasis/potential priority):

Slido Results: Respond: *Priorities (1 of 4)*

RESPOND: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (1/4)

0 2 7

Incident Management (RS.MA)

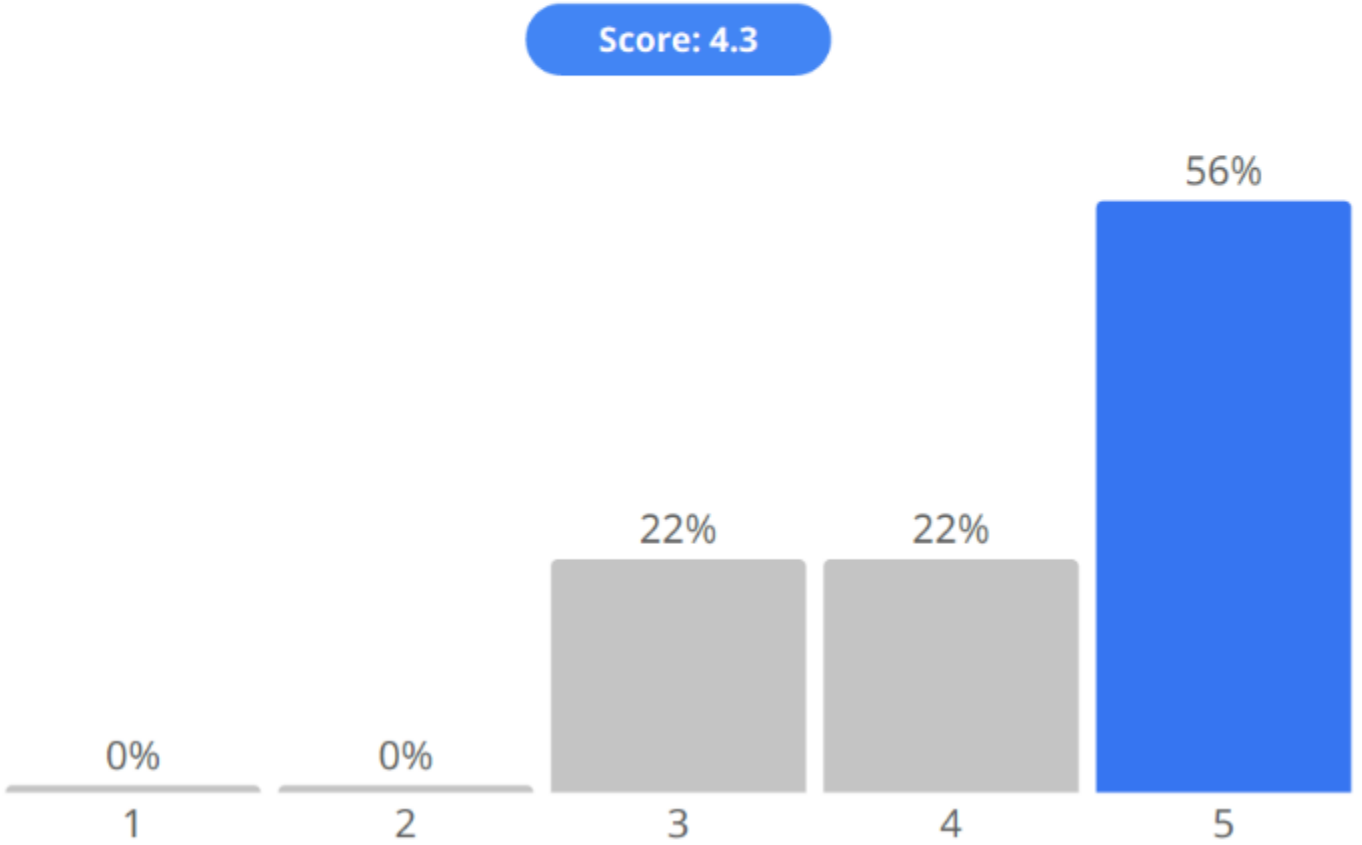


Slido Results: Respond: *Priorities (2 of 4)*

RESPOND: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (2/4)

0 2 7

Incident Analysis (RS.AN)

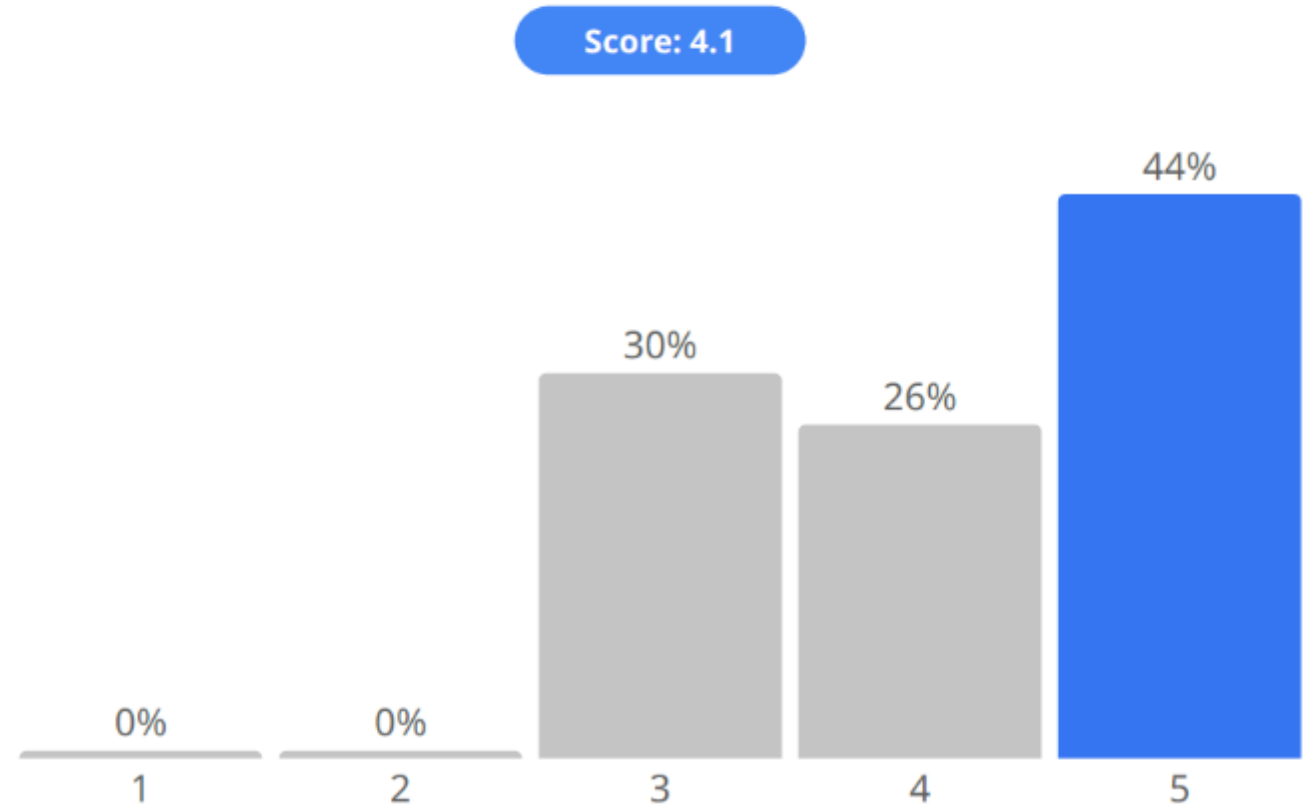


Slido Results: Respond: *Priorities (3 of 4)*

RESPOND: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (3/4)

0 2 7

Incident Response Reporting and Communication (RS.CO)



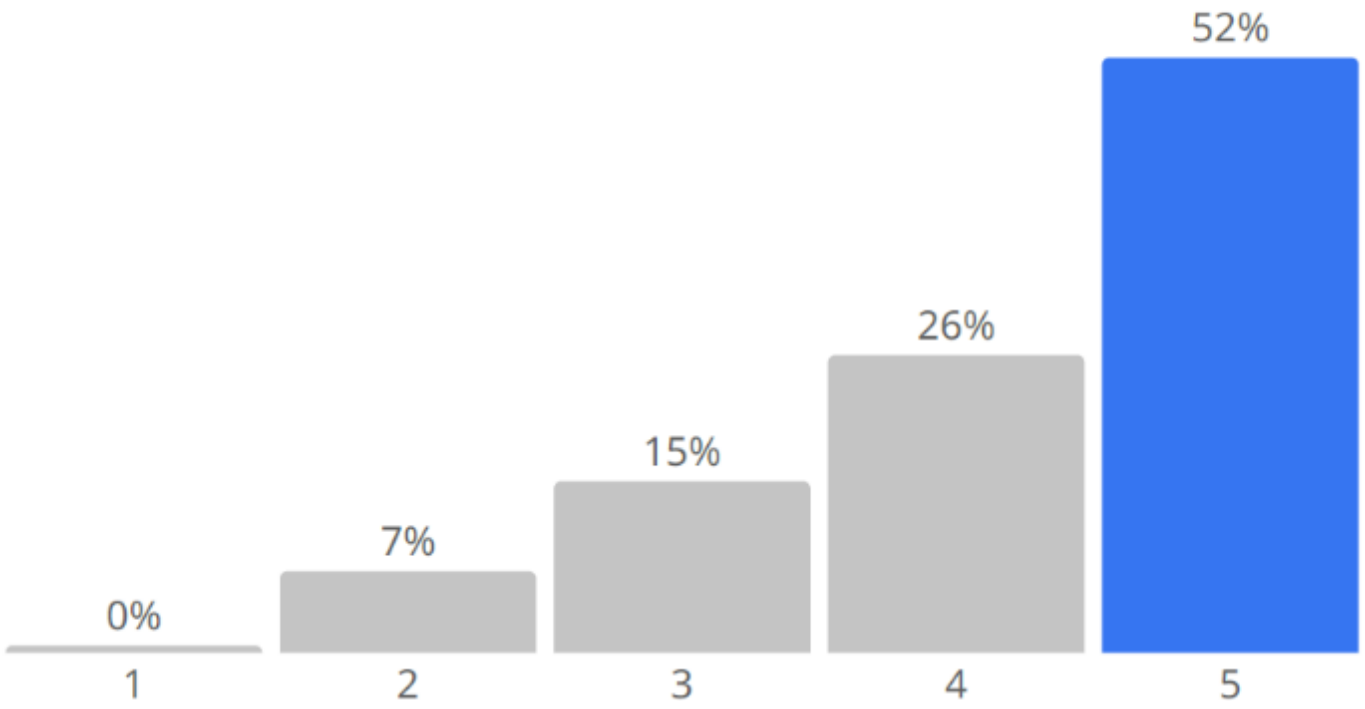
Slido Results: Respond: *Priorities (4 of 4)*

RESPOND: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (4/4)

0 2 7

Incident Mitigation (RS.MI)

Score: 4.2



Recover: Opportunities

Category	Description
Incident Recovery Plan Execution (RC.RP)	Restoration activities are performed to ensure operational availability of systems and services affected by cybersecurity incidents.
Incident Recovery Communications (RC.CO)	Restoration activities are coordinated with internal and external parties.

Recover: Priorities

Slido.com
#CyberAI_WS

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Category	Description	Heatmap
Incident Recovery Plan Execution (RC.RP)	Restoration activities are performed to ensure operational availability of systems and services affected by cybersecurity incidents.	0.3
Incident Recovery Communications (RC.CO)	Restoration activities are coordinated with internal and external parties.	0.1

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of emphasis/potential priority):

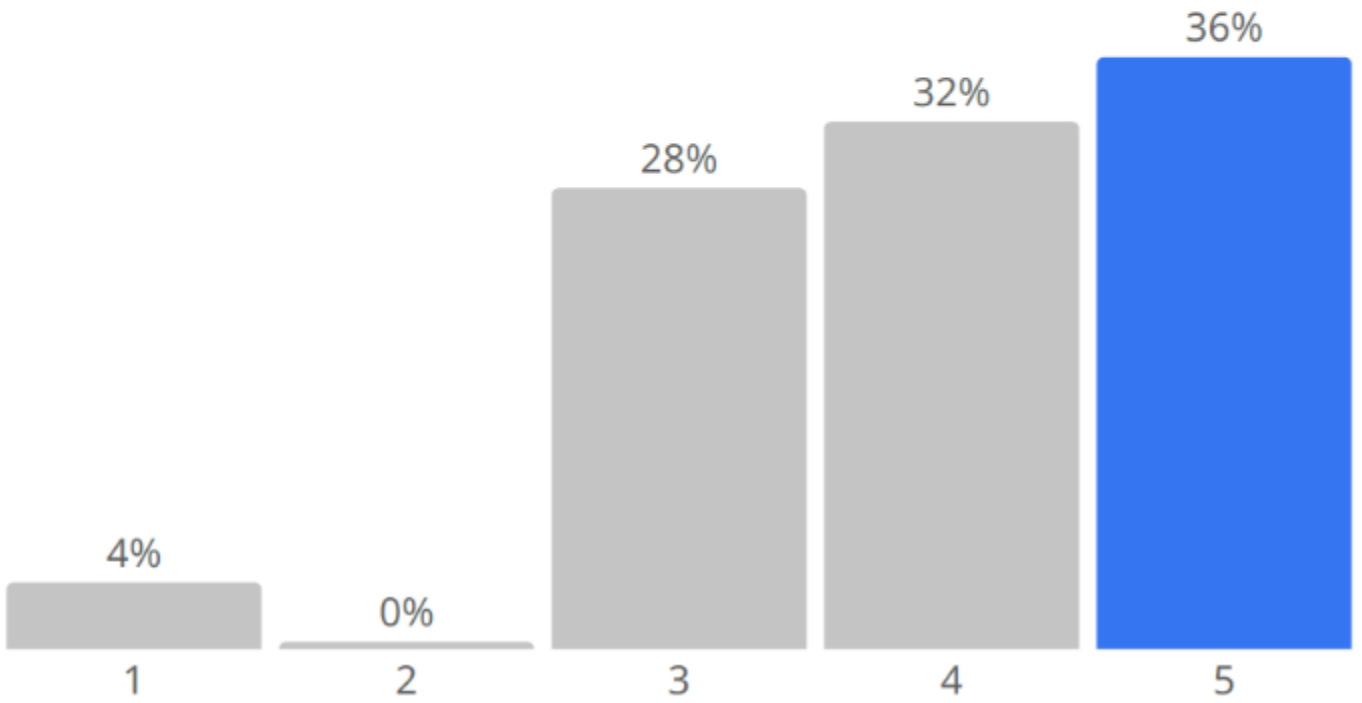
Slido Results: *Recover: Priorities (1 of 2)*

RECOVER: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (1/2)

0 2 5

Incident Recovery Plan Execution (RC.RP)

Score: 4.0

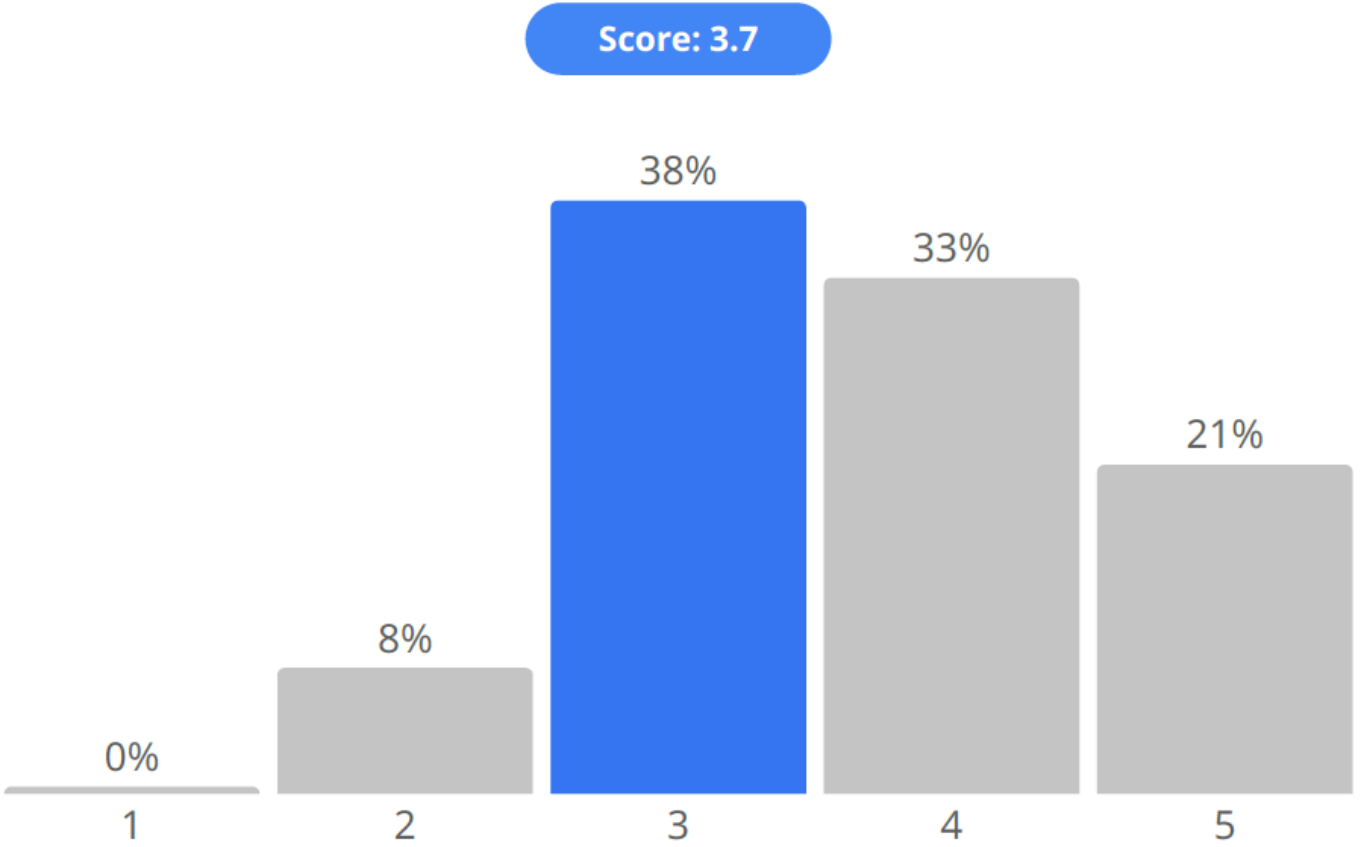


Slido Results: *Recover: Priorities (2 of 2)*

RECOVER: How important are these Categories to Conducting AI-enabled Cyber Defense? (1 = Not Important, 2 = Somewhat Important, 3 = Important, 4 = Very Important, 5 = Extremely Important) (2/2)

0 2 4

Incident Recovery Communication (RC.CO)



Close-out

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

We Appreciate Your Input

THANK YOU

Your input is a critical part of this process! Thank you for contributing to the development of the Cyber AI Profile!

Slido.com
#CyberAI_WS

Slido Results: *Close-out (1 of 3)*

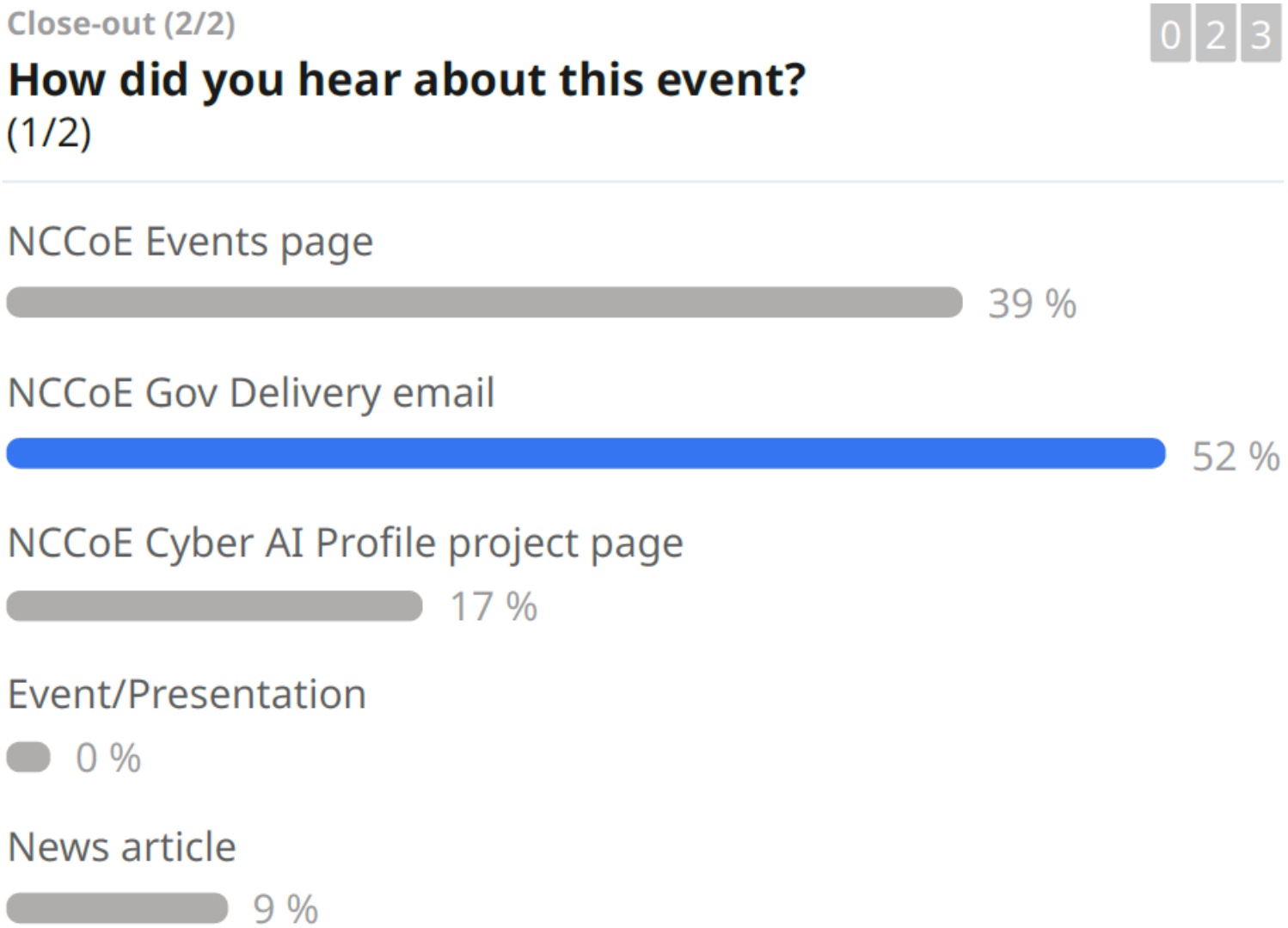
Close-out (1/2)

005

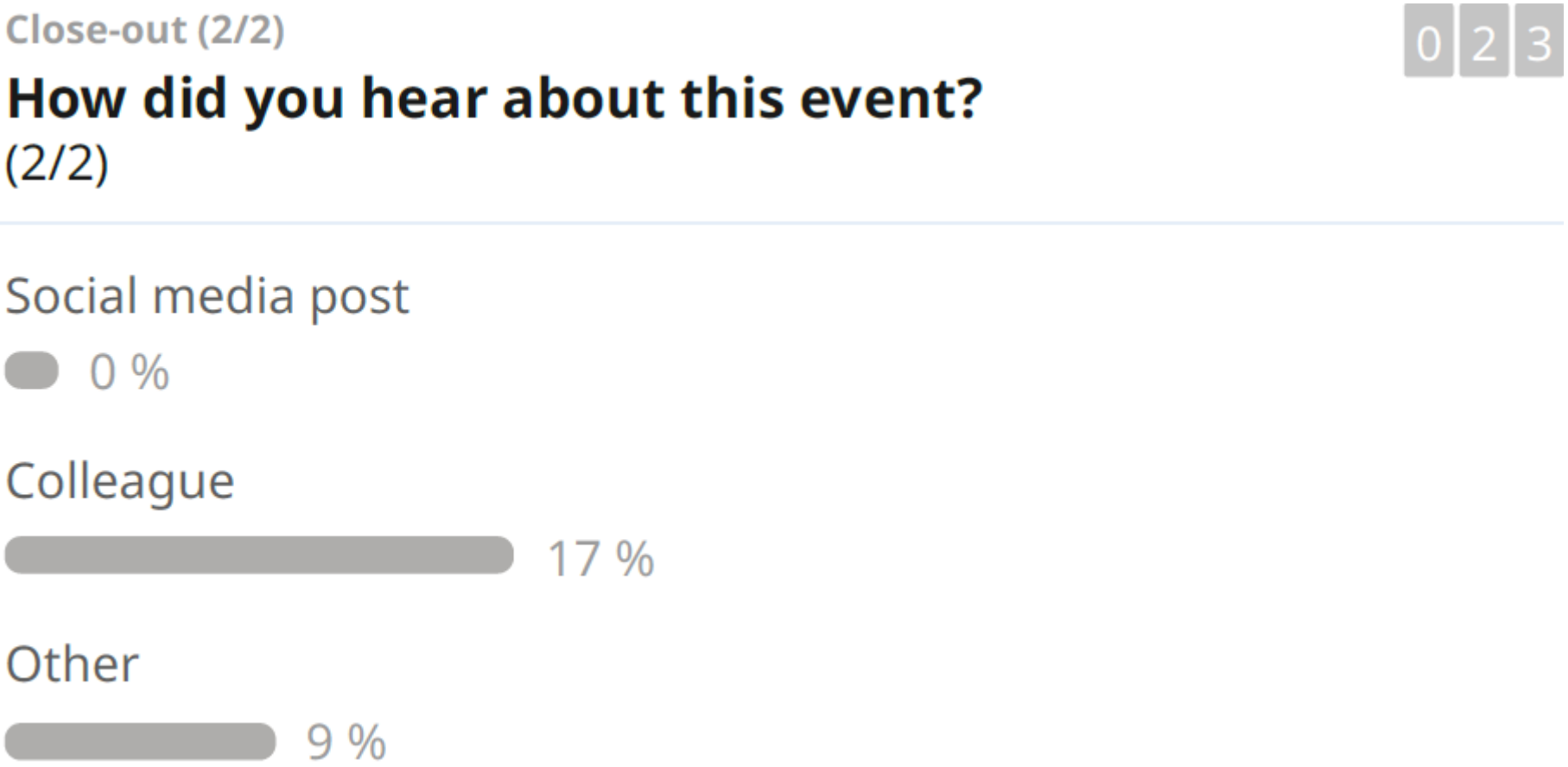
What additional resources should we incorporate into our research?

- 😀
- It was adequate.
- CSA AI Controls Matrix
- TBD
- None

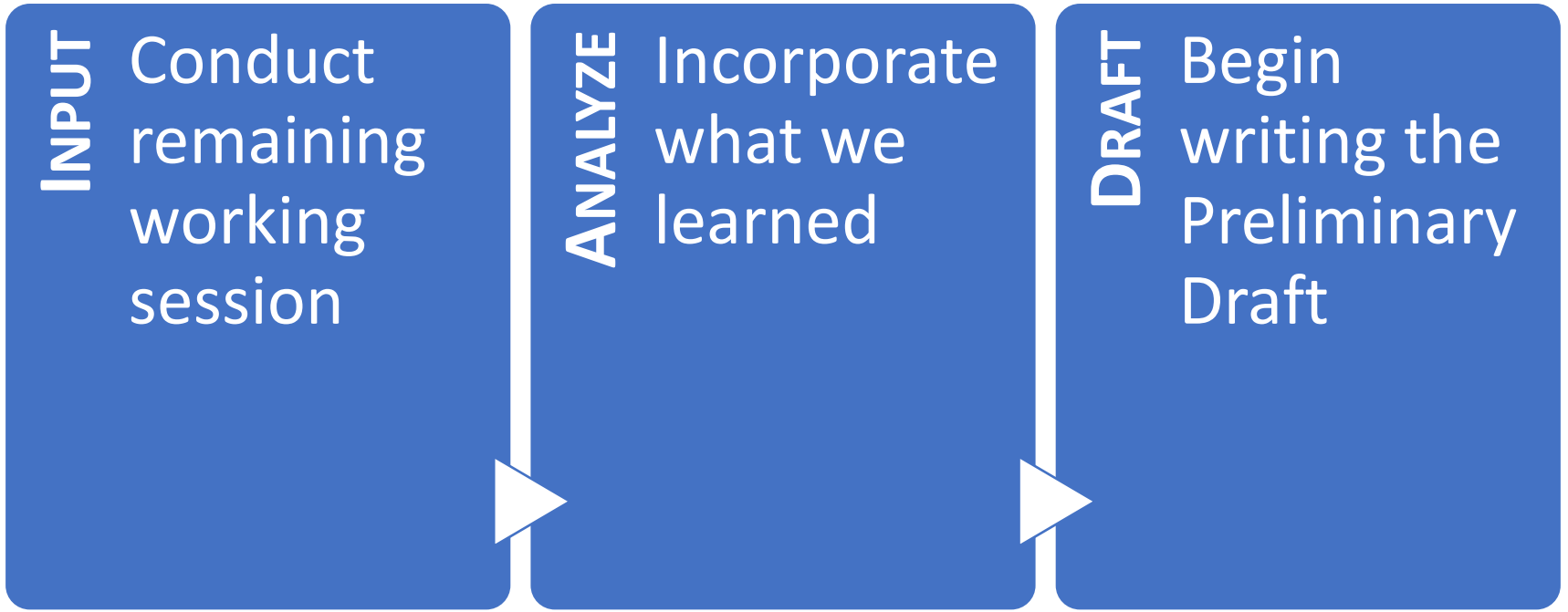
Slido Results: *Close-out (2 of 3)*



Slido Results: *Close-out (3 of 3)*



Working Sessions Next Steps



If you have a resource we should review during our analysis or we missed your input today, please feel free to email us: CyberAIProfile@nist.gov!

Working Session Schedule

August 19, 2025

Conducting AI-enabled Cyber Defense

☑

August 26, 2025

Securing AI System Components

September 2, 2025

Thwarting AI-enabled Cyber Attacks

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Resources

Cyber AI Profile

- NIST Cybersecurity, Privacy, and AI Program
- Blog post: Managing Cybersecurity and Privacy Risks in the Age of Artificial Intelligence: Launching a New Program at NIST | NIST
- NCCoE Project Page: Cyber AI Profile
- Cybersecurity and AI Workshop Concept Paper (posted in advance of the April 3, 2025, workshop)
- April 3rd Cyber AI Profile Workshop recording
- Blog post: Reflections from the First Cyber AI Profile Workshop
- Cyber AI Profile COI Working Sessions Introduction Video

NIST Cybersecurity Framework

- https://www.nist.gov/cyberframework/
- https://www.nist.gov/cyberframework/faqs
- https://www.nist.gov/informative-references
- https://www.nist.gov/cyberframework/events-and-presentations

NIST Resources for Applying NIST Frameworks

- https://www.nccoe.nist.gov/applying-frameworks-resources

Community Profiles

- https://www.nccoe.nist.gov/examples-community-profiles
- https://www.nccoe.nist.gov/creating-community-profiles-faqs

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

https://www.nccoe.nist.gov/projects/cyber-ai-profile

CyberAIProfile@nist.gov

nccoe.nist.gov

@NISTcyber