Cyber AI Profile Working Sessions: Securing AI System Components

August 26, 2025

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

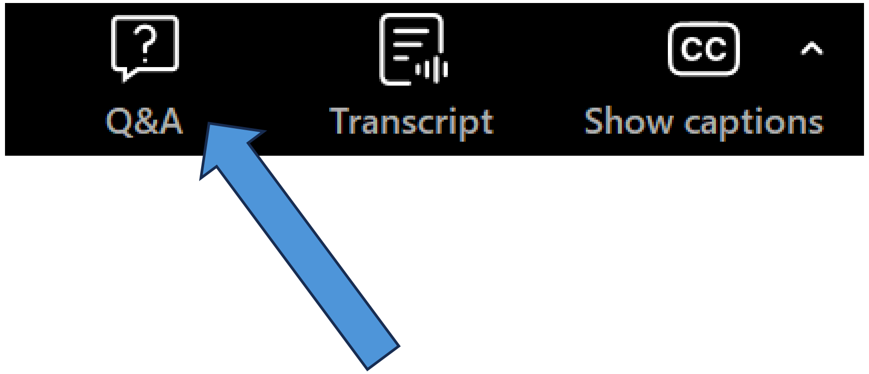
Agenda

- Cyber AI Profile Project Overview
- Today’s Plan
- Refresher from Working Session Introduction
- CSF 2.0 Category Considerations: Securing AI System Components
- Close-out

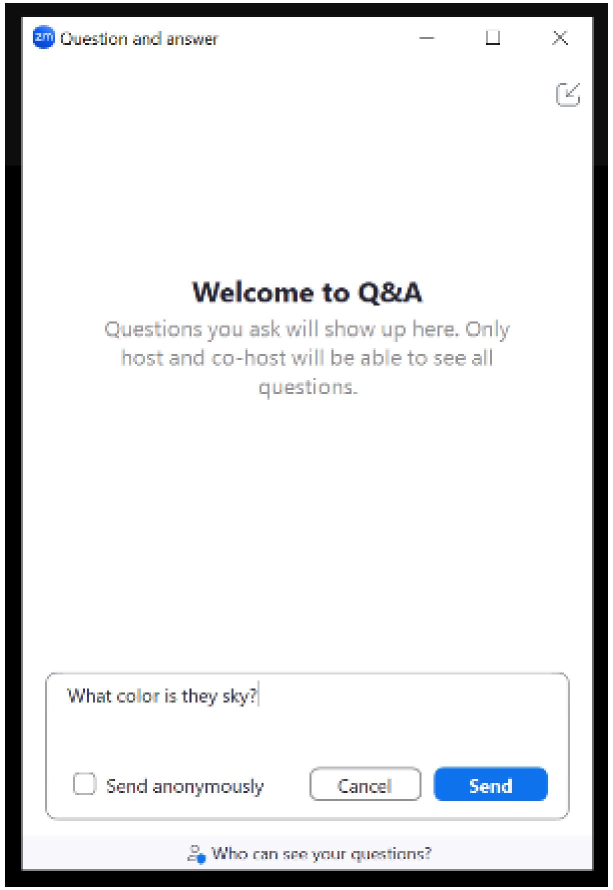
Submitting Questions

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



Welcome to Q&A

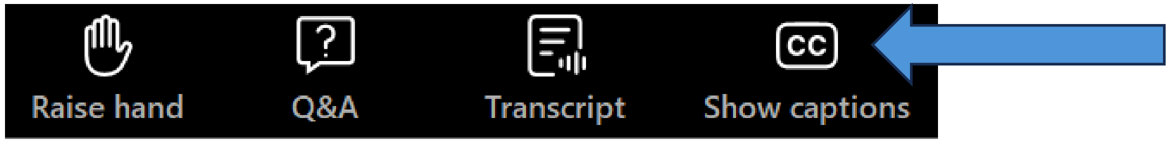
Questions you ask will show up here. Only host and co-host will be able to see all questions.

What color is they sky?

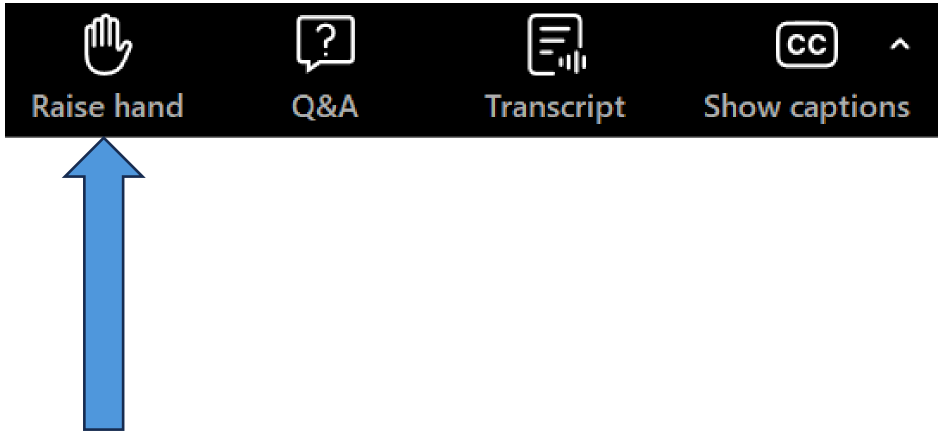
2. Type your question in the text box and click Send

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

Purpose:

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

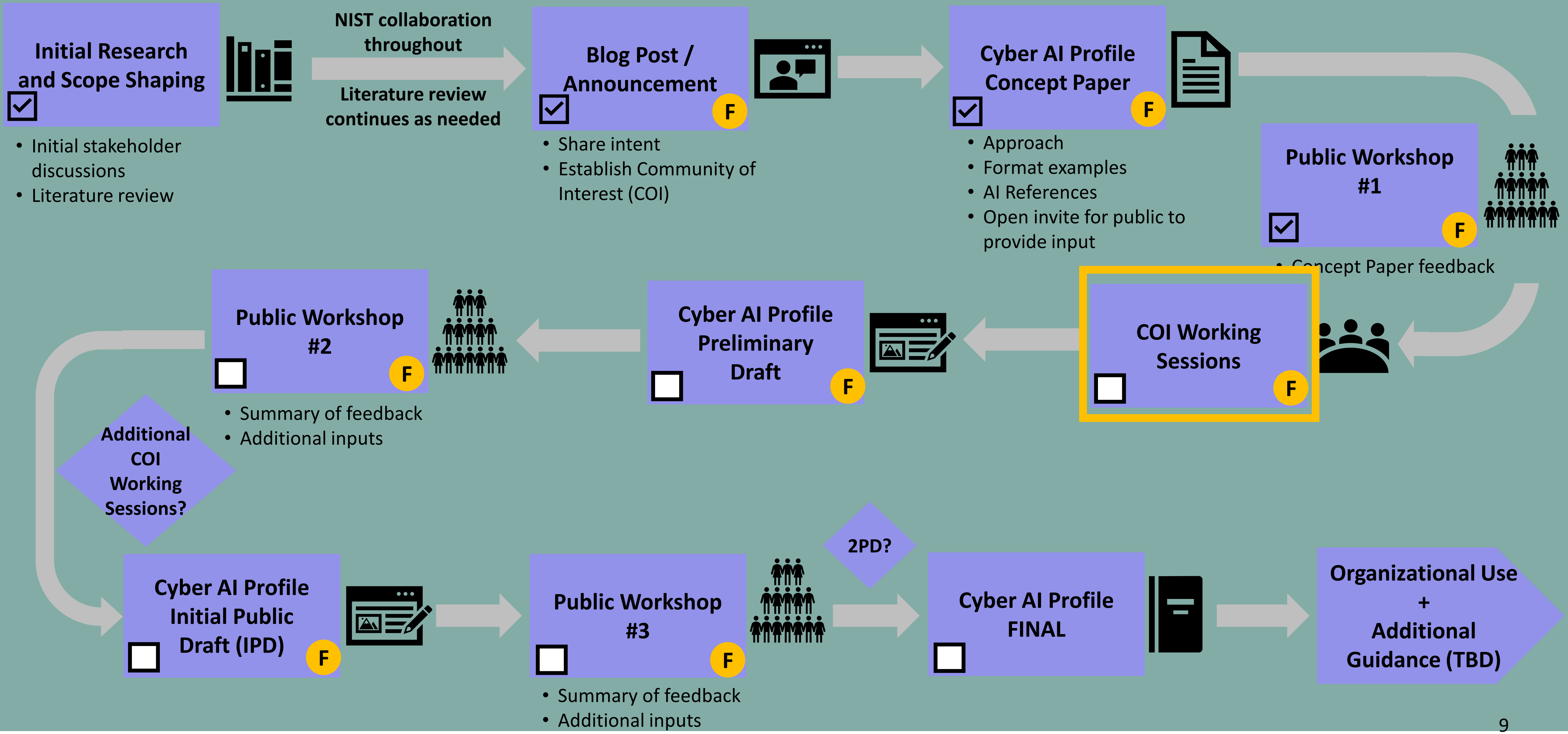
Areas of focus:

- 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



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)

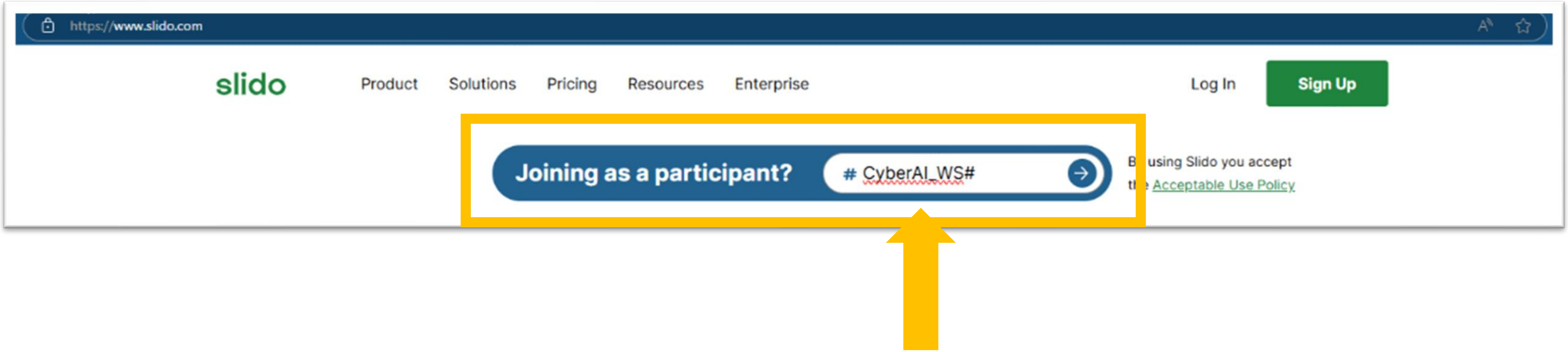
Today’s Plan

How You Contribute Today

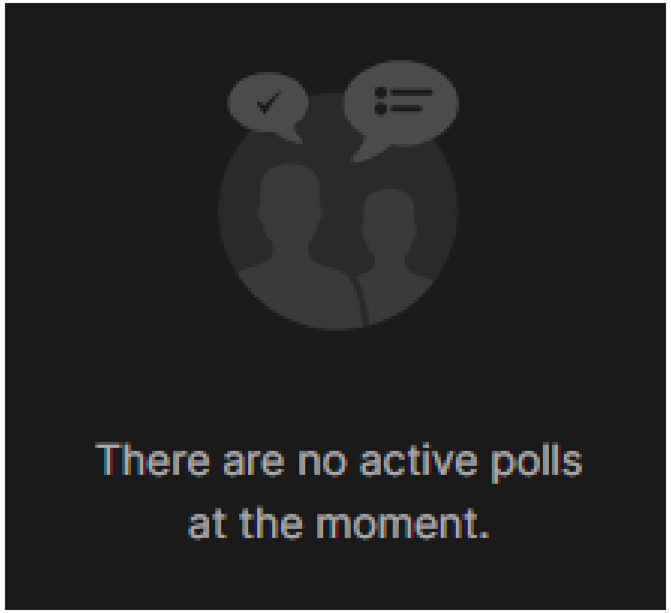
- **Please raise your virtual hand or type in the chat to contribute**
- Members of the press, please identify yourself and your organization
- Be respectful of others
- Please don't be shy – we would love to hear from everyone!
- Please remain on mute when not speaking
- We will use Slido to facilitate some of our discussions

Using Slido

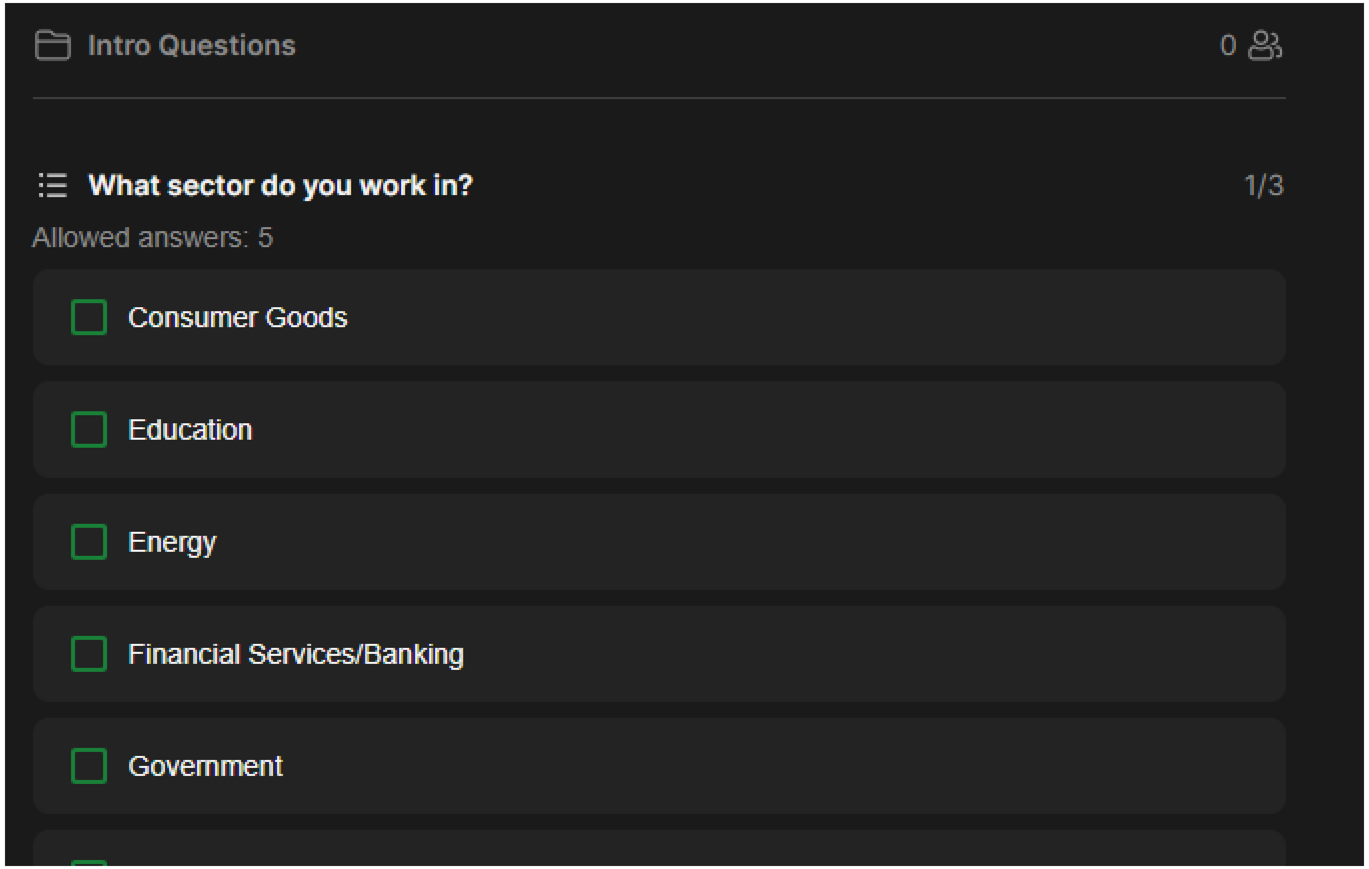
- We will be using Slido to facilitate some of our discussions
- Options to join via QR code or URL + event code
- Works on mobile phone and computer
- Responses are anonymous
- We want to hear your feedback – this provides a useful way to capture what we hear from you



Slido Screens



There are no active polls at the moment.



Intro Questions 0

What sector do you work in? 1/3

Allowed answers: 5

- □ Consumer Goods
- □ Education
- □ Energy
- □ Financial Services/Banking
- □ Government

Slido: Getting to Know You

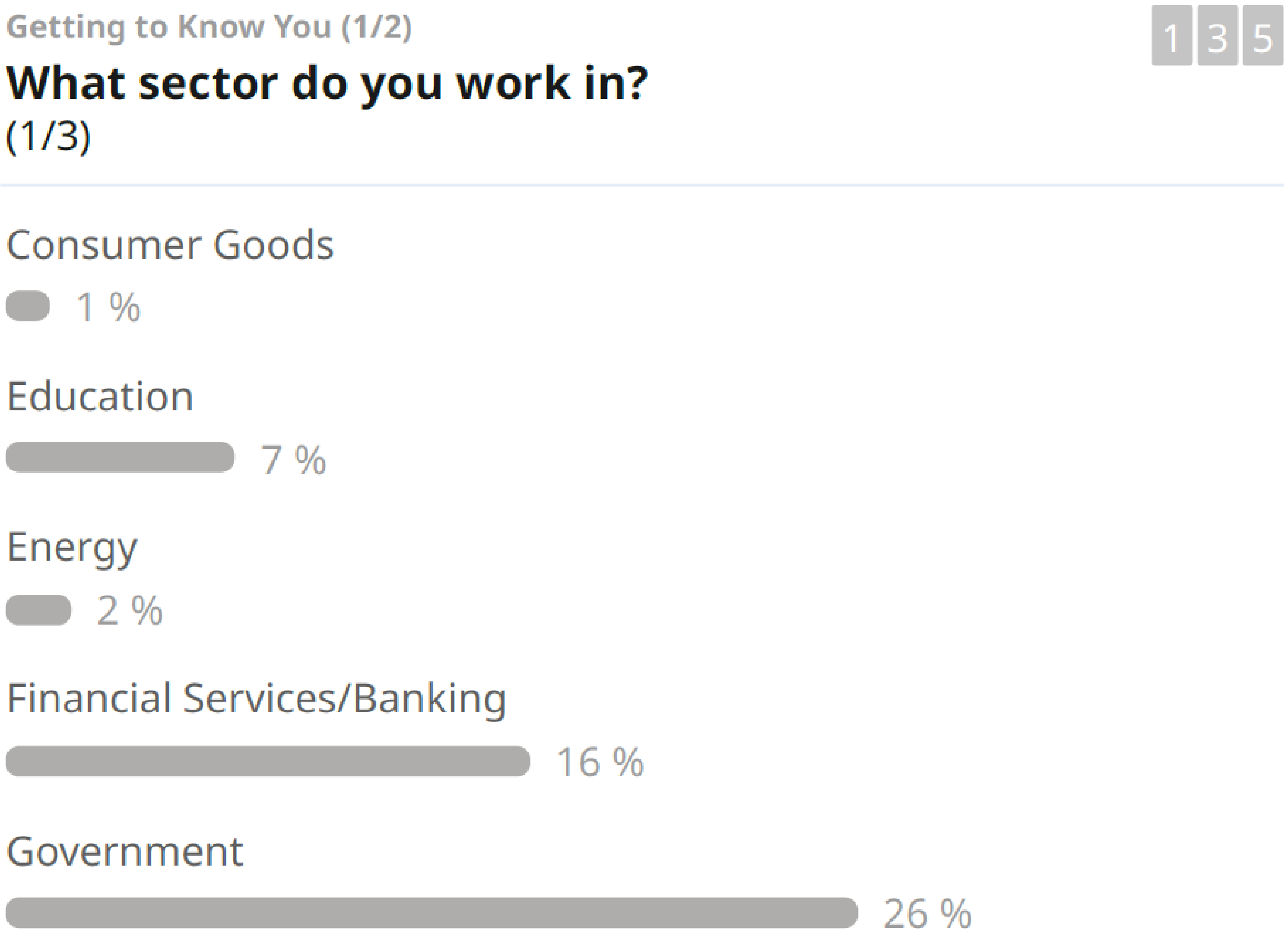
- What sector do you work in?
- Which NIST Frameworks does your organization use?

Join at

Slido.com

#CyberAI_WS1

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



Getting to Know You (1/2)

1 3 5

What sector do you work in?
(1/3)

Consumer Goods

1 %

Education

7 %

Energy

2 %

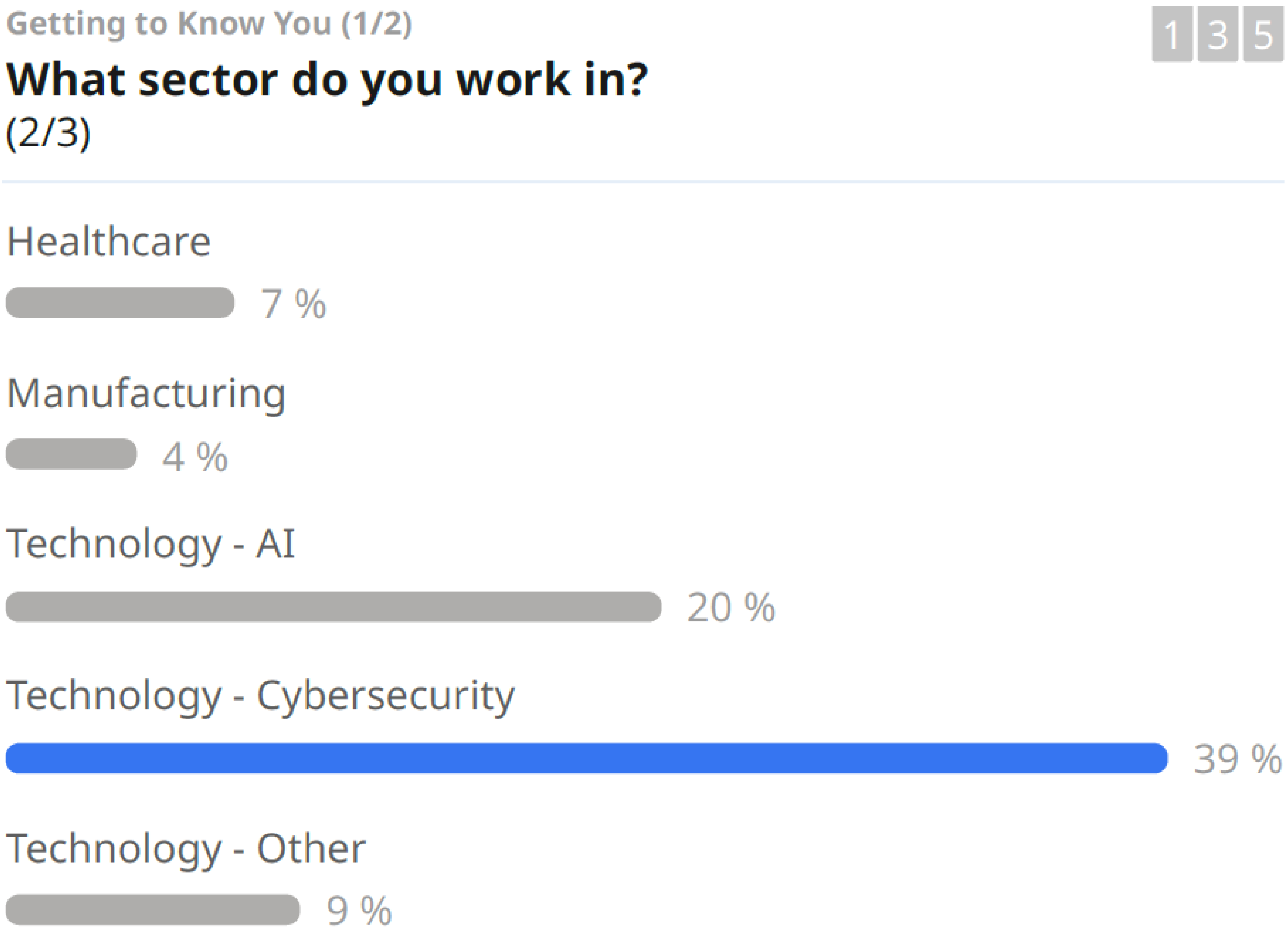
Financial Services/Banking

16 %

Government

26 %

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



Getting to Know You (1/2)

What sector do you work in?
(2/3)

Healthcare
7 %

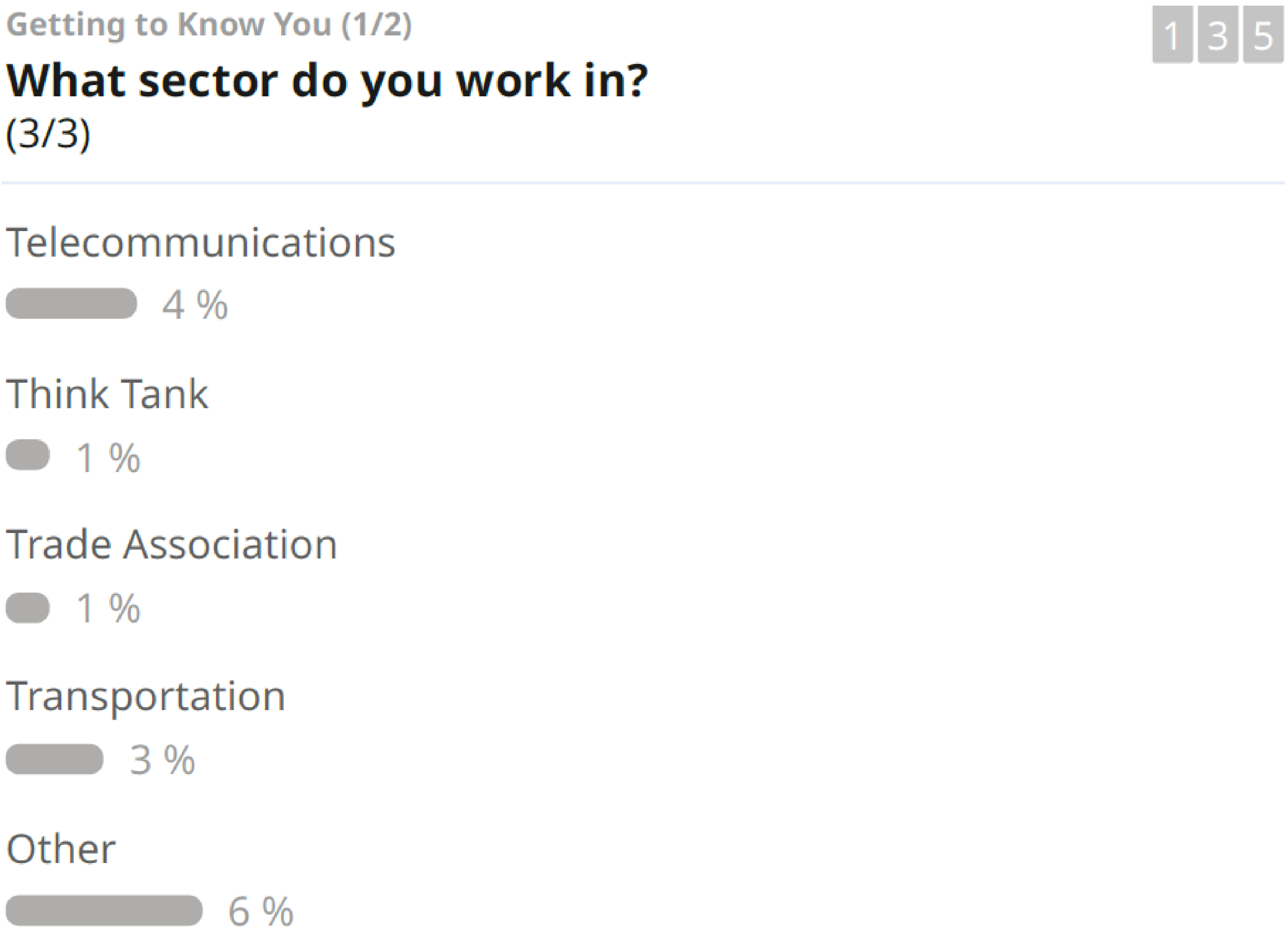
Manufacturing
4 %

Technology - AI
20 %

Technology - Cybersecurity
39 %

Technology - Other
9 %

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



Getting to Know You (1/2)

135

What sector do you work in?
(3/3)

Telecommunications

4 %

Think Tank

1 %

Trade Association

1 %

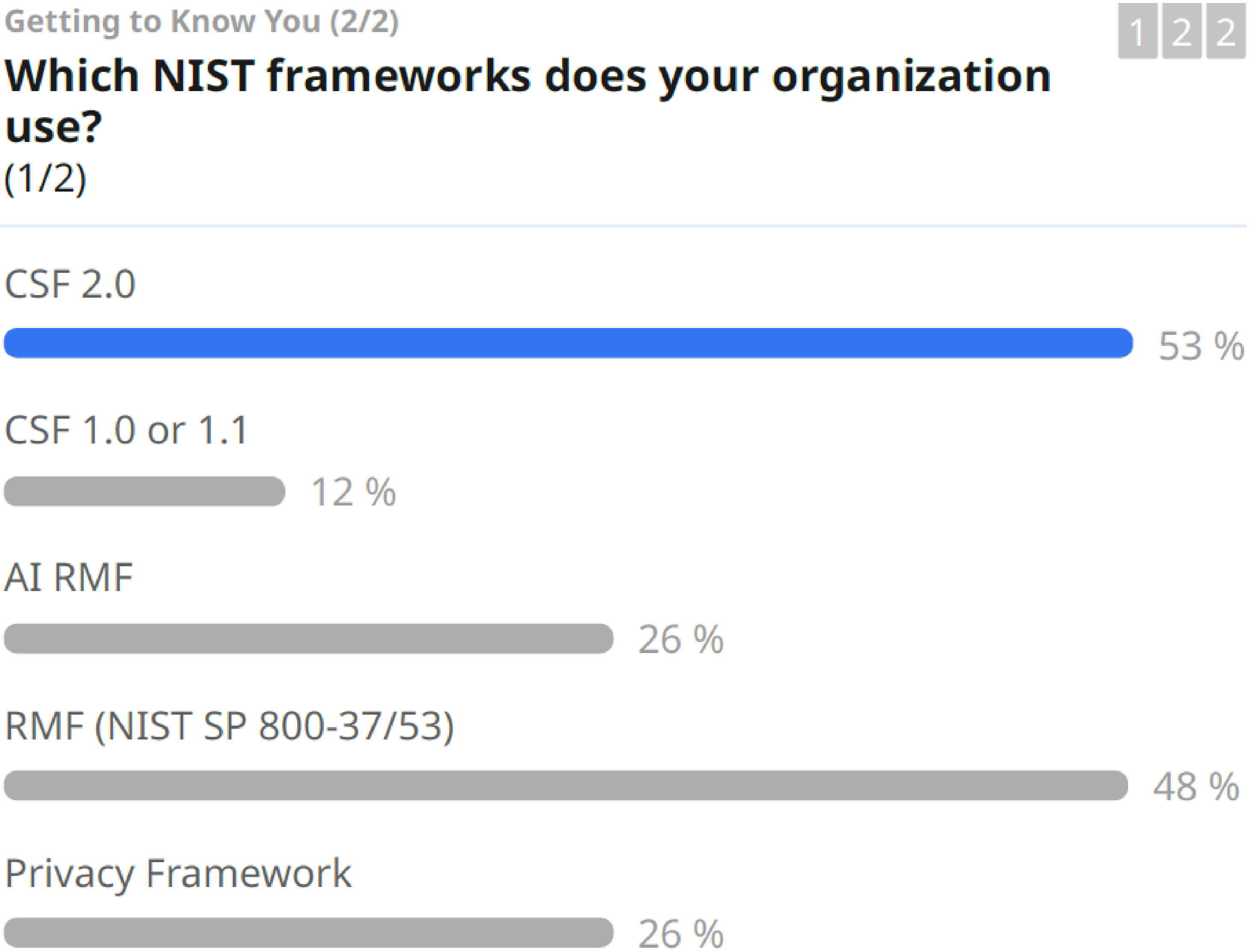
Transportation

3 %

Other

6 %

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



Getting to Know You (2/2)

1 2 2

Which NIST frameworks does your organization use?
(1/2)

CSF 2.0 53 %

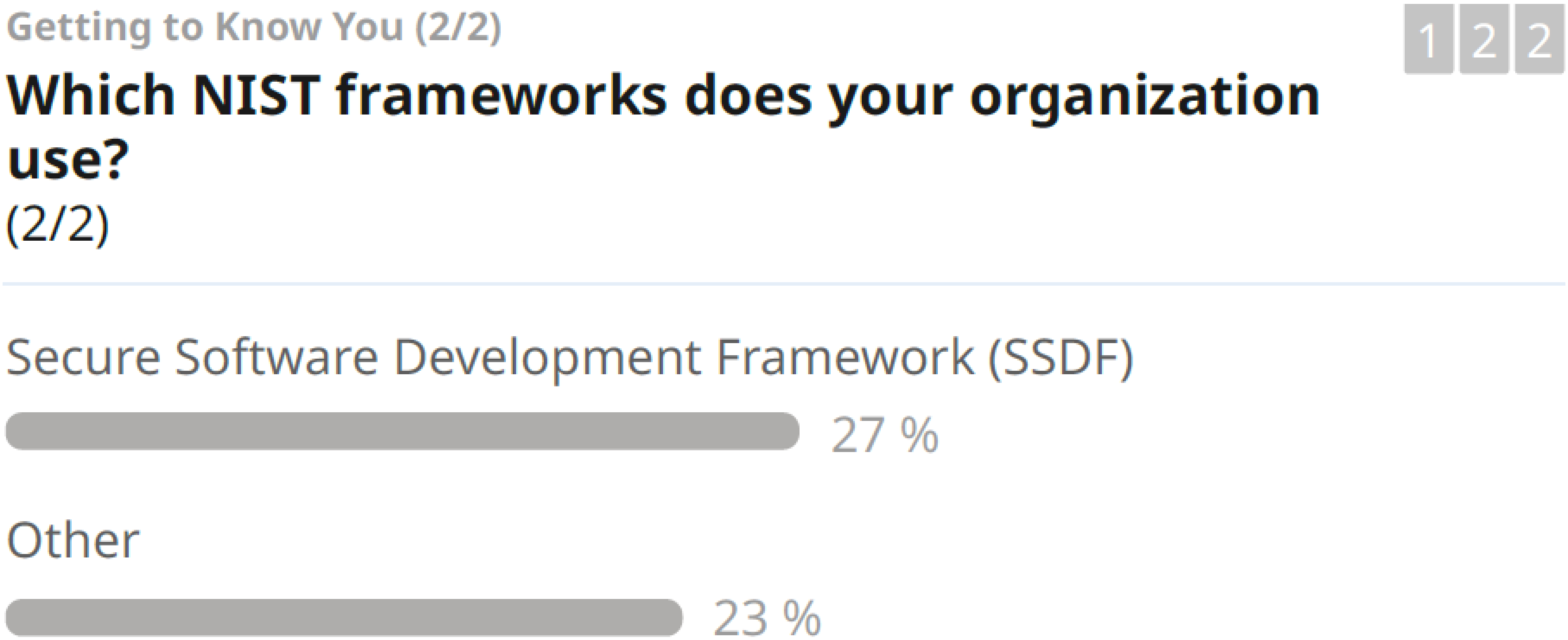
CSF 1.0 or 1.1 12 %

AI RMF 26 %

RMF (NIST SP 800-37/53) 48 %

Privacy Framework 26 %

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



Getting to Know You (2/2)

Which NIST frameworks does your organization use?
(2/2)

Secure Software Development Framework (SSDF)

27 %

Other

23 %

Today’s Focus: CSF 2.0 Categories for Securing AI System Components

- **Scope:** Identifying cybersecurity challenges when integrating AI into organizational ecosystems and infrastructure
- **Focus Area characteristics – cybersecurity considerations regarding:**
  - Preparing an organization to develop and/or deploy AI capabilities
  - Managing and AI system throughout is lifecycle
  - Understanding the threat and vulnerability landscape for AI technology components
- **Examples of AI Technology System Components:**
  - AI Model (e.g., LLMs, computer vision models, RL)
  - Data pipeline and storage (e.g., understanding provenance, training, cleaning, anomalies)
  - System Interfaces, internal tool callings (e.g., AI agents, interfaces)
  - Third-party AI platforms/services, external tool callings (e.g., LLMs, AI agents, APIs)

Challenge Areas

- Data Governance, Security, and Privacy
- Adversarial Attacks
- Unauthorized Access and Use
- Identifying, Detecting, and Responding to AI Component Vulnerabilities and Adverse Events
- Supply Chain Security
- Model Drift and Onset of Unexpected or Inaccurate Results

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:
    - What is unique about securing AI systems?
    - 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)?

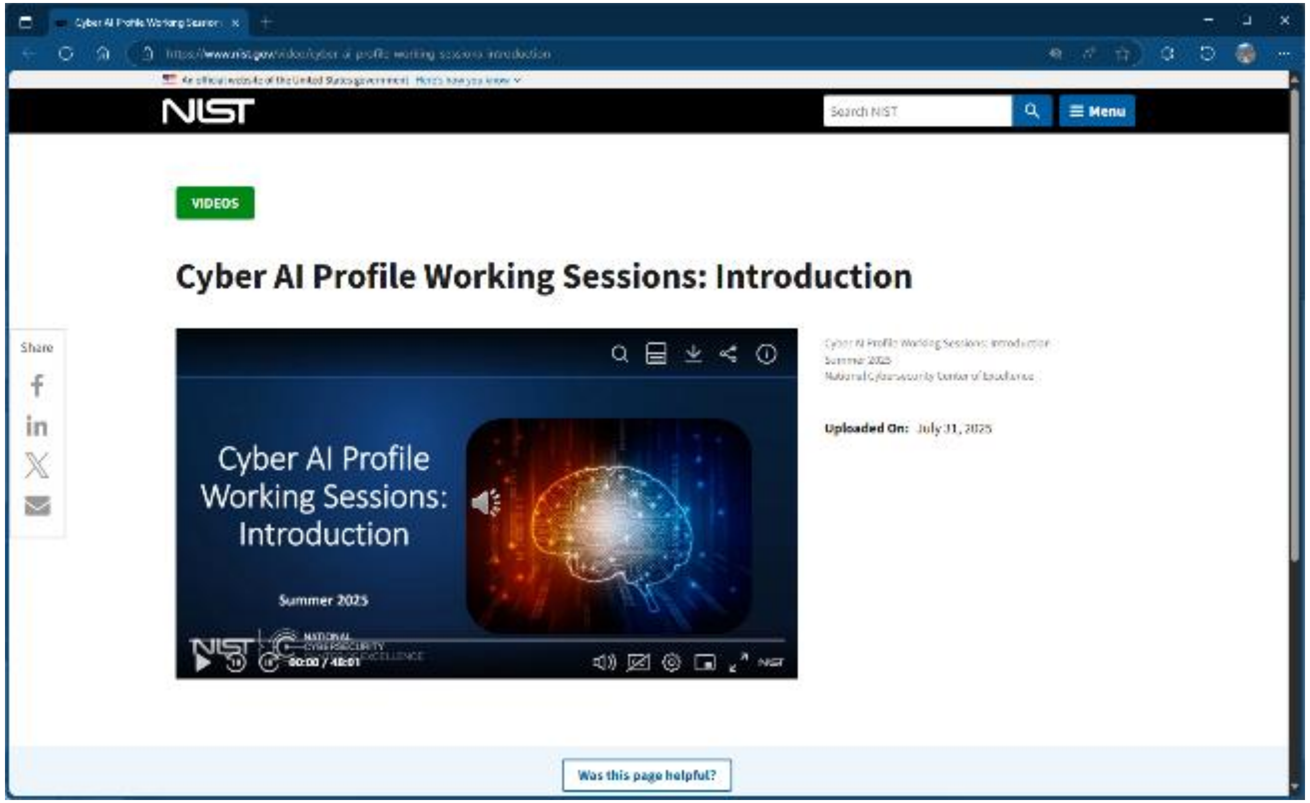
CSF 2.0 PDF

Cybersecurity Framework page

Refresher from Working Session Introduction

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Working Sessions Introduction Video


Cyber AI Profile Working Session
https://www.nist.gov/video/cyber-ai-profile-working-sessions-introduction
An official website of the United States government Here's how you know
NIST
Search NIST
Menu
Share
Cyber AI Profile Working Sessions: Introduction
Summer 2025
00:00 / 48:01
Was this page helpful?

VIDEOS

Cyber AI Profile Working Sessions: Introduction

Cyber AI Profile Working Sessions: Introduction
Summer 2025
National Cybersecurity Center of Excellence

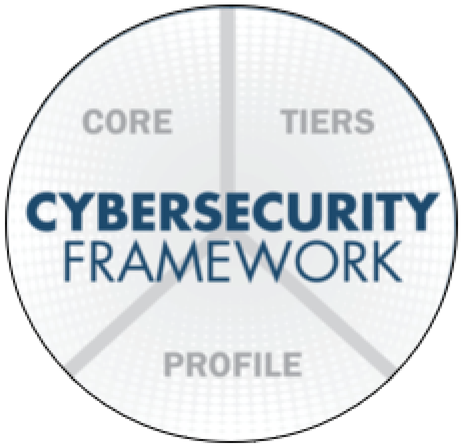
Uploaded On: July 31, 2025

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



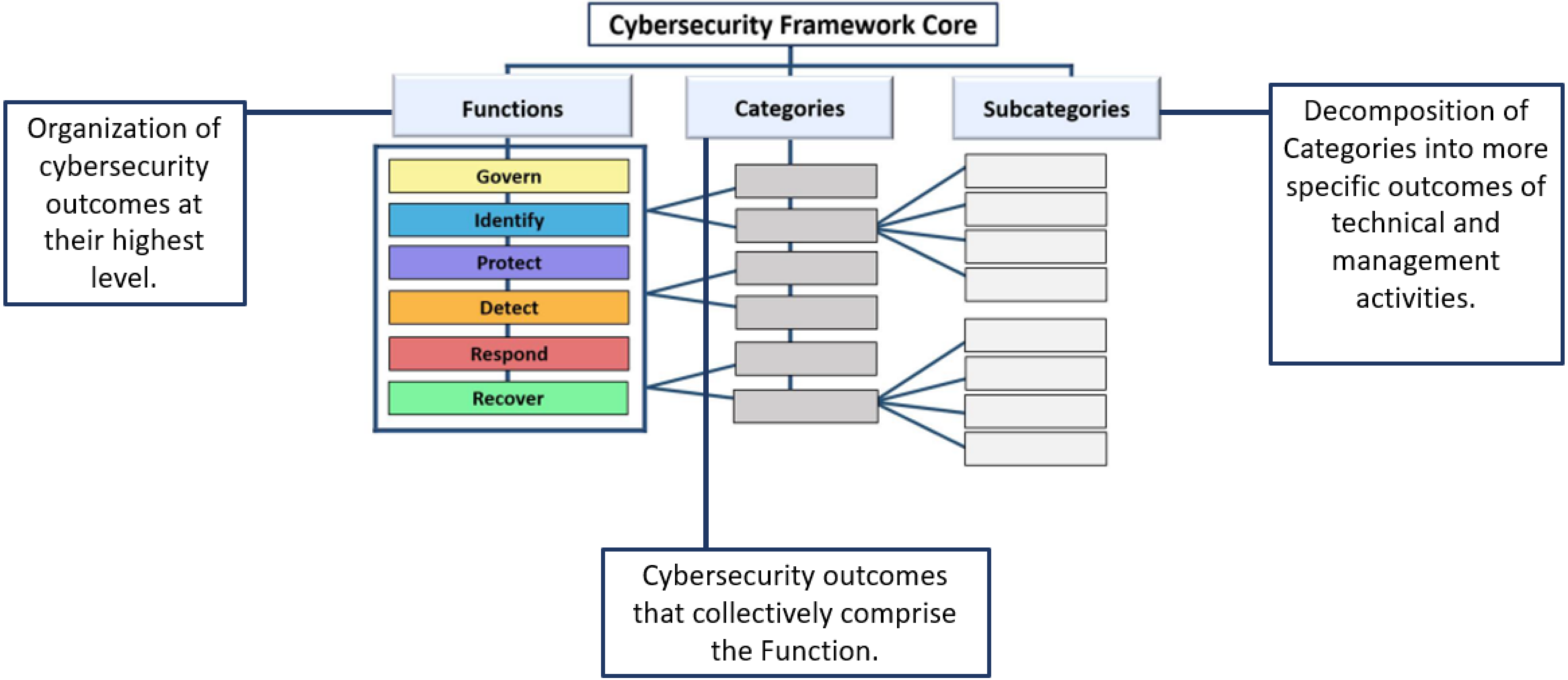
Help characterize the context and rigor of an organization’s cybersecurity risk governance and management practices

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



Cybersecurity Framework Core

Functions

Categories

Subcategories

Govern

Identify

Protect

Detect

Respond

Recover

Organization of cybersecurity outcomes at their highest level.

Decomposition of Categories into more specific outcomes of technical and management activities.

Cybersecurity outcomes that collectively comprise the Function.

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.]

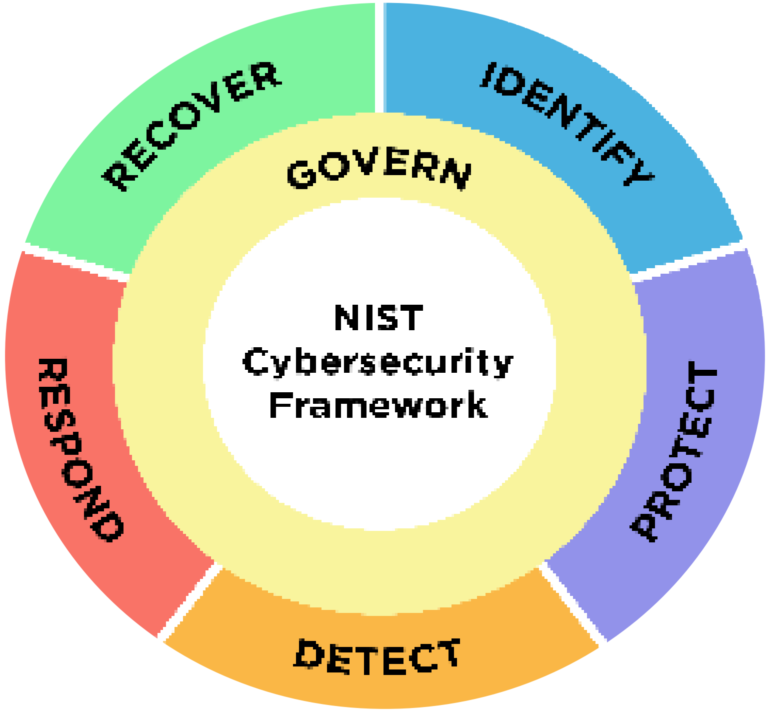
Example Content - Extreme Fast Charging Profile (CSF 1.1)

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: Securing AI System Components

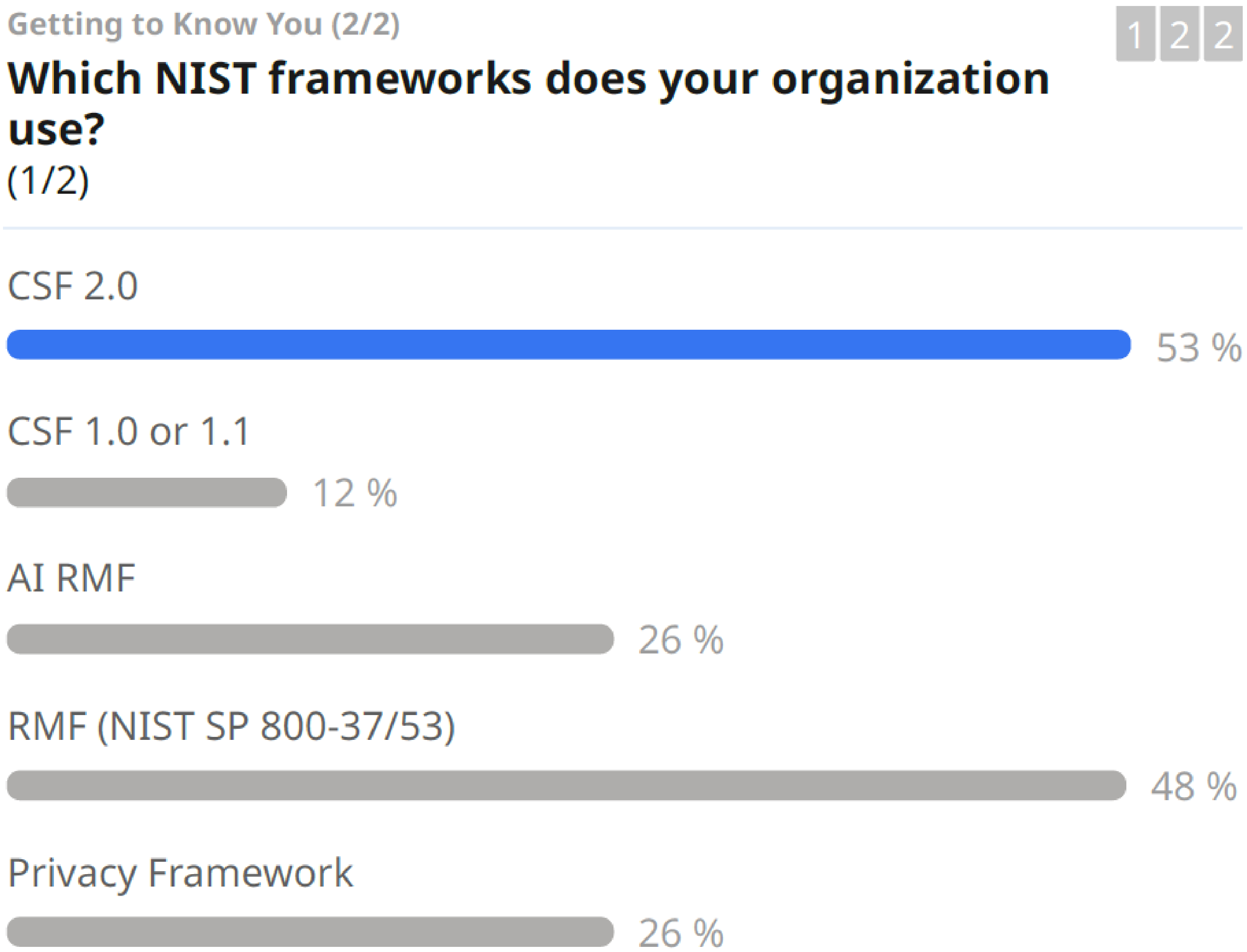
Importance of CSF Functions for Securing AI System Components

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE



Slido.com
#CyberAI_WS1

Slido Results: *Importance of CSF Functions for Securing AI System Components (1 of 2)*



Getting to Know You (2/2)

1 2 2

Which NIST frameworks does your organization use?
(1/2)

CSF 2.0

53 %

CSF 1.0 or 1.1

12 %

AI RMF

26 %

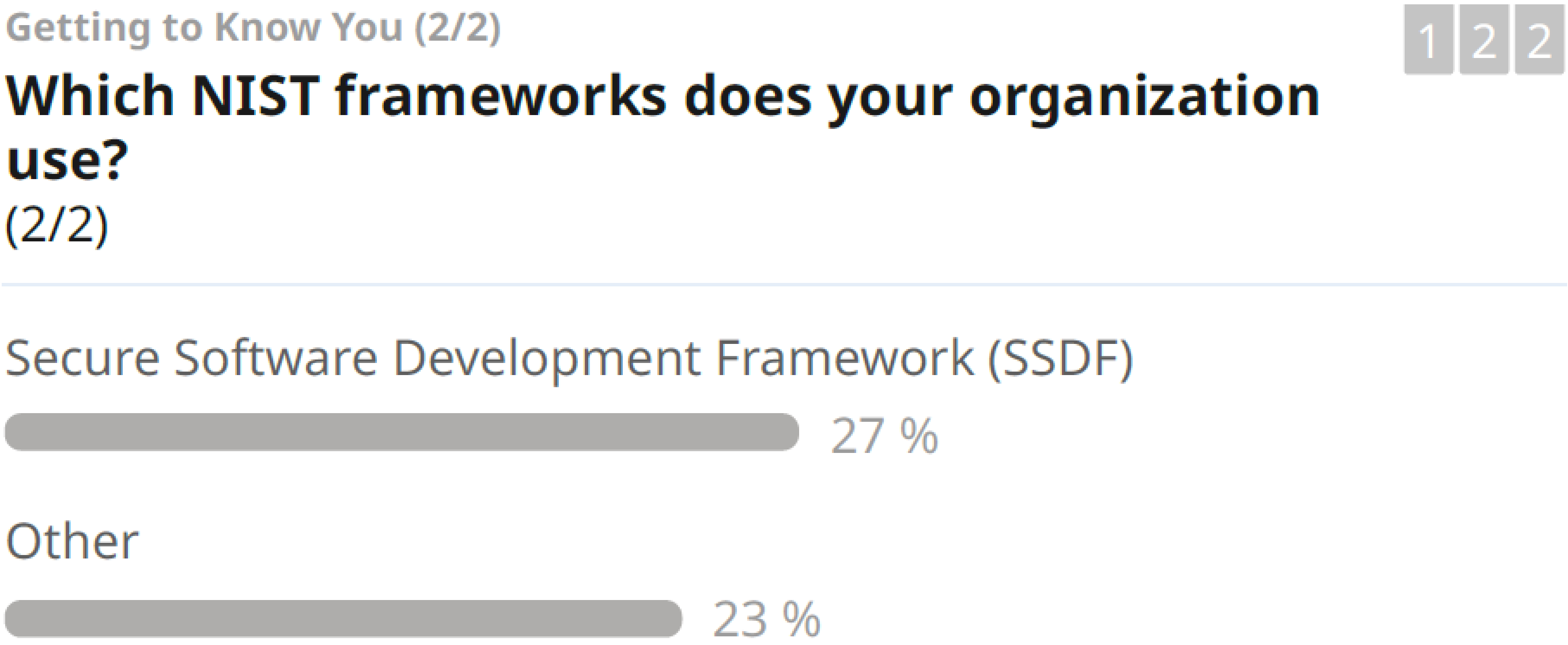
RMF (NIST SP 800-37/53)

48 %

Privacy Framework

26 %

Slido Results: *Importance of CSF Functions for Securing AI System Components (2 of 2)*



Getting to Know You (2/2)

Which NIST frameworks does your organization use?
(2/2)

Secure Software Development Framework (SSDF)
27 %

Other
23 %

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
- **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.

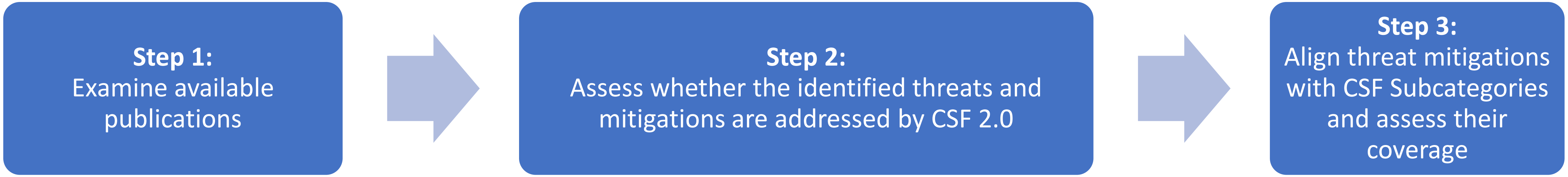
Example Sources of Inputs

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 we use in our analysis?
- Are there critical mitigations that are missing from the current body of work?

Align Industry Mitigations to NIST CSF 2.0



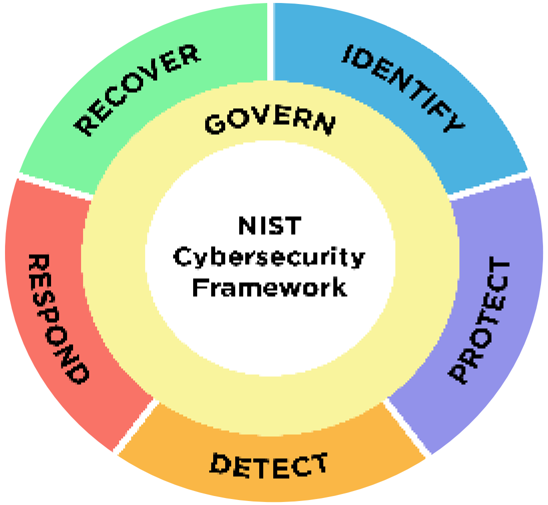
Step 1:
Examine available publications

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^{TM} - OWASP



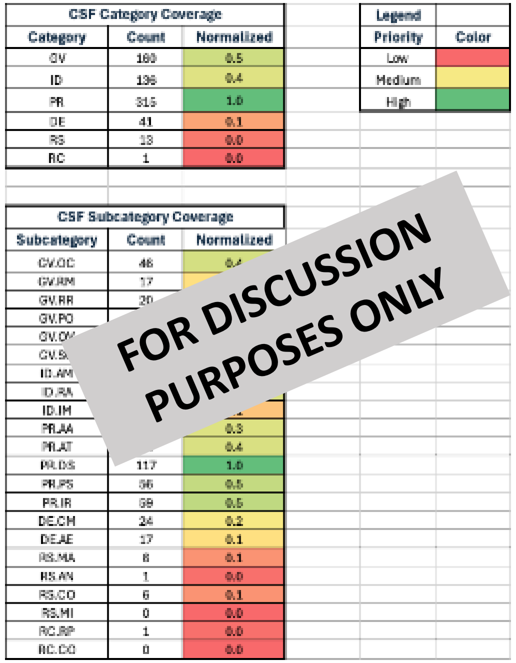
NIST Special Publication 800-53
Revision 5

Security and Privacy Controls for
Information Systems and Organizations

[illegible]

[illegible]

NIST
[illegible]



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	[illegible]
GV.RR	20	[illegible]
GV.PO	[illegible]	[illegible]
GV.OV	[illegible]	[illegible]
GV.SC	[illegible]	[illegible]
ID.AM	[illegible]	[illegible]
ID.RA	[illegible]	[illegible]
ID.IM	[illegible]	0.1
PR.AA	[illegible]	0.3
PR.AT	[illegible]	0.4
PR.DS	117	1.0
PR.PS	56	0.5
PR.IR	59	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

Summary View

GOVERN	Heatmap
Organizational Context (GV.OC)	0.40
Risk Management Strategy (GV.RM)	0.56
Roles, Responsibilities, and Authorities (GV.RR)	0.06
Policy (GV.PO)	0.12
Oversight (GV.OV)	0.27
Cybersecurity Supply Chain Risk Management (GV.SC)	1.0

IDENTIFY	Heatmap
Asset Management (ID.AM)	0.65
Risk Assessment (ID.RA)	0.92
Improvement (ID.IM)	0.44

PROTECT	Heatmap
Identity Management, Authentication and Access Control (PR.AA)	0.42
Awareness and Training (PR.AT)	0.10
Data Security (PR.DS)	0.37
Platform Security (PR.PS)	0.44
Technology Infrastructure Resilience (PR.IR)	0.31

DETECT	Heatmap
Continuous Monitoring (DE.CM)	0.56
Adverse Event Analysis (DE.AE)	0.48

RESPOND	Heatmap
Incident Management (RS.MA)	0.19
Incident Analysis (RS.AN)	0.15
Incident Response Reporting and Communication (RS.CO)	0.17
Incident Mitigation (RS.MI)	0.02

RECOVER	Heatmap
Incident Recovery Plan Execution (RC.RP)	0.13
Incident Recovery Communications (RC.CO)	0.10

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of current emphasis): Low Moderate High

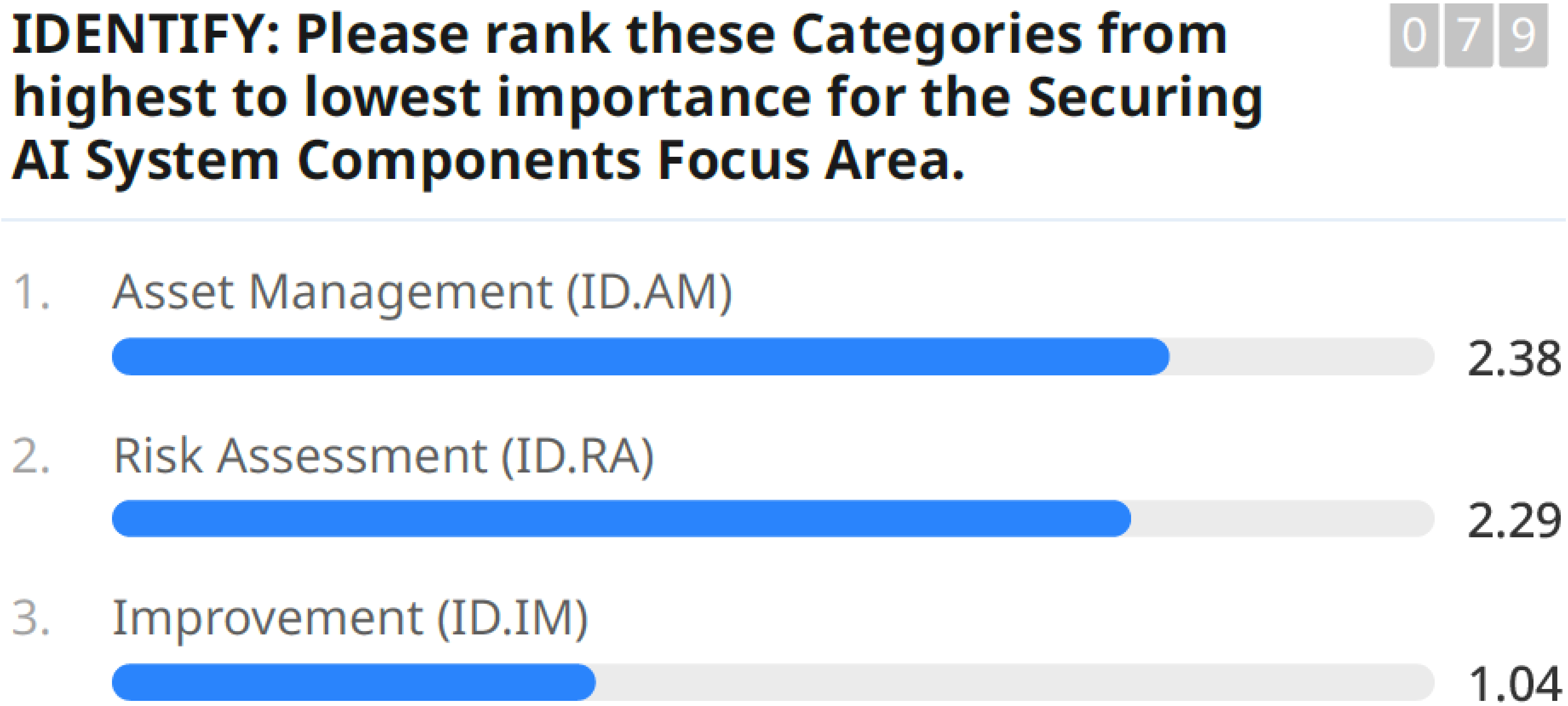
Identify

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.	0.65
Improvement (ID.IM)	Improvements to organizational cybersecurity risk management processes, procedures and activities are identified across all CSF Functions.	0.44
Risk Assessment (ID.RA)	The cybersecurity risk to the organization, assets, and individuals is understood by the organization.	0.92

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of current emphasis): **Low** — **Moderate** — **High**

Slide Results: *Identify (1 of 2)*

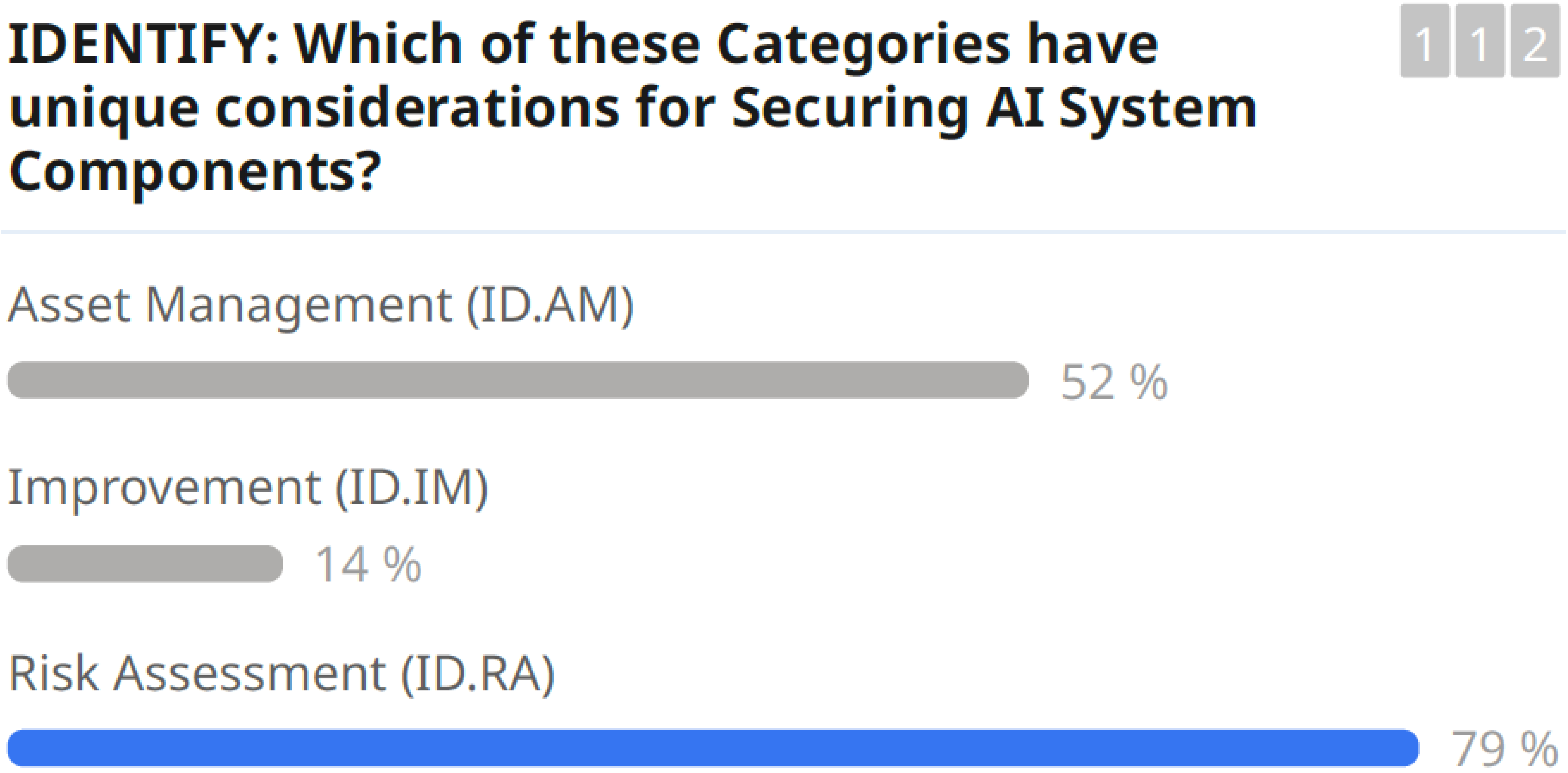


IDENTIFY: Please rank these Categories from highest to lowest importance for the Securing AI System Components Focus Area.

079

1. Asset Management (ID.AM) 2.38
2. Risk Assessment (ID.RA) 2.29
3. Improvement (ID.IM) 1.04

Slide Results: *Identify (2 of 2)*



IDENTIFY: Which of these Categories have unique considerations for Securing AI System Components?

Detect

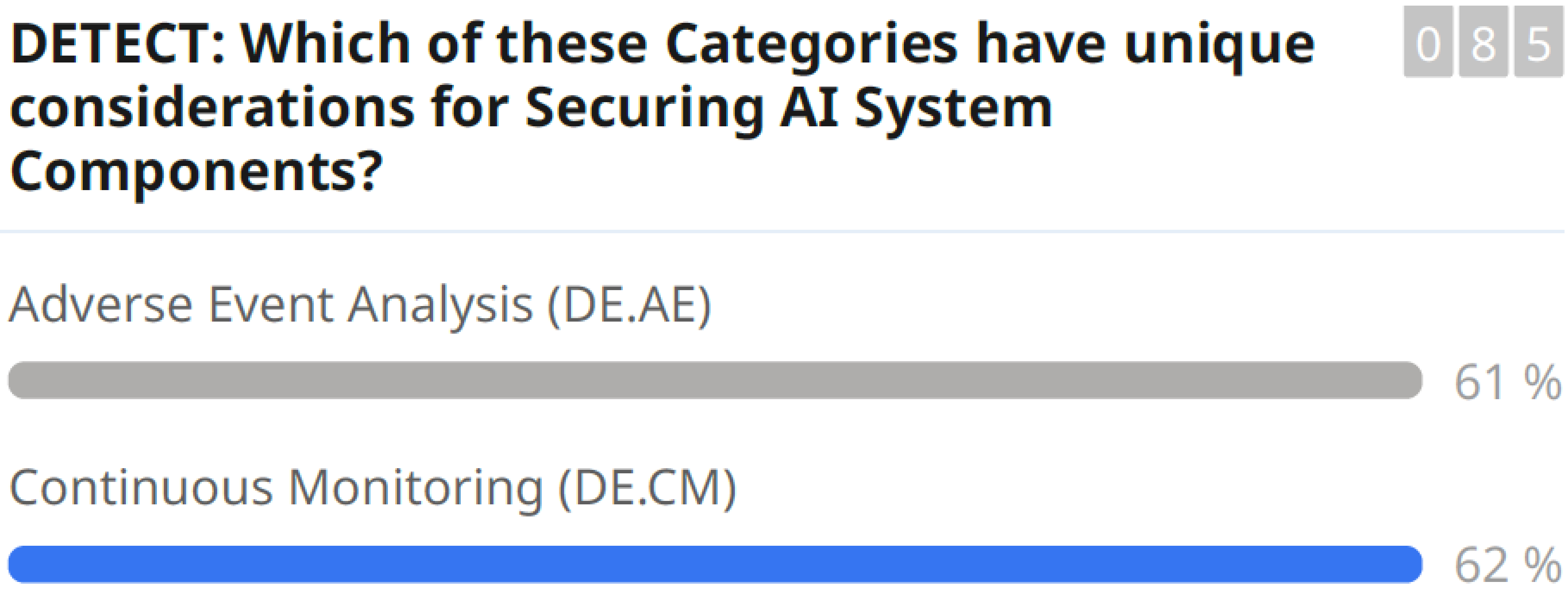
Category	Description	Heatmap
Adverse Event Analysis (DE.AE)	Anomalies, indicators of compromise, and other potentially adverse events are analyzed to characterize the events and detect cybersecurity incidents.	0.48
Continuous Monitoring (DE.CM)	Assets are monitored to find anomalies, indicators of compromise, and other potentially adverse events.	0.56

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of current emphasis):

Low **Moderate** **High**

Slido Results: *Detect (1 of 2)*



DETECT: Which of these Categories have unique considerations for Securing AI System Components?

0 8 5

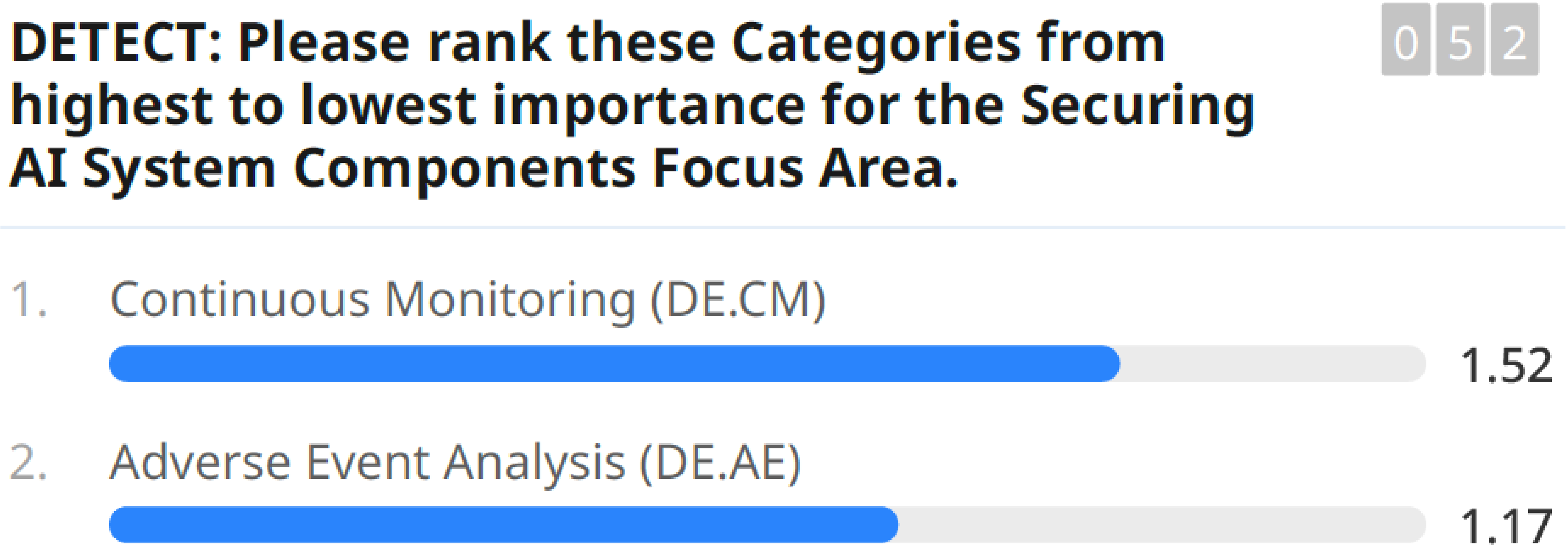
Adverse Event Analysis (DE.AE)

61 %

Continuous Monitoring (DE.CM)

62 %

Slido Results: *Detect (2 of 2)*



0 5 2

DETECT: Please rank these Categories from highest to lowest importance for the Securing AI System Components Focus Area.

1. Continuous Monitoring (DE.CM) 1.52
2. Adverse Event Analysis (DE.AE) 1.17

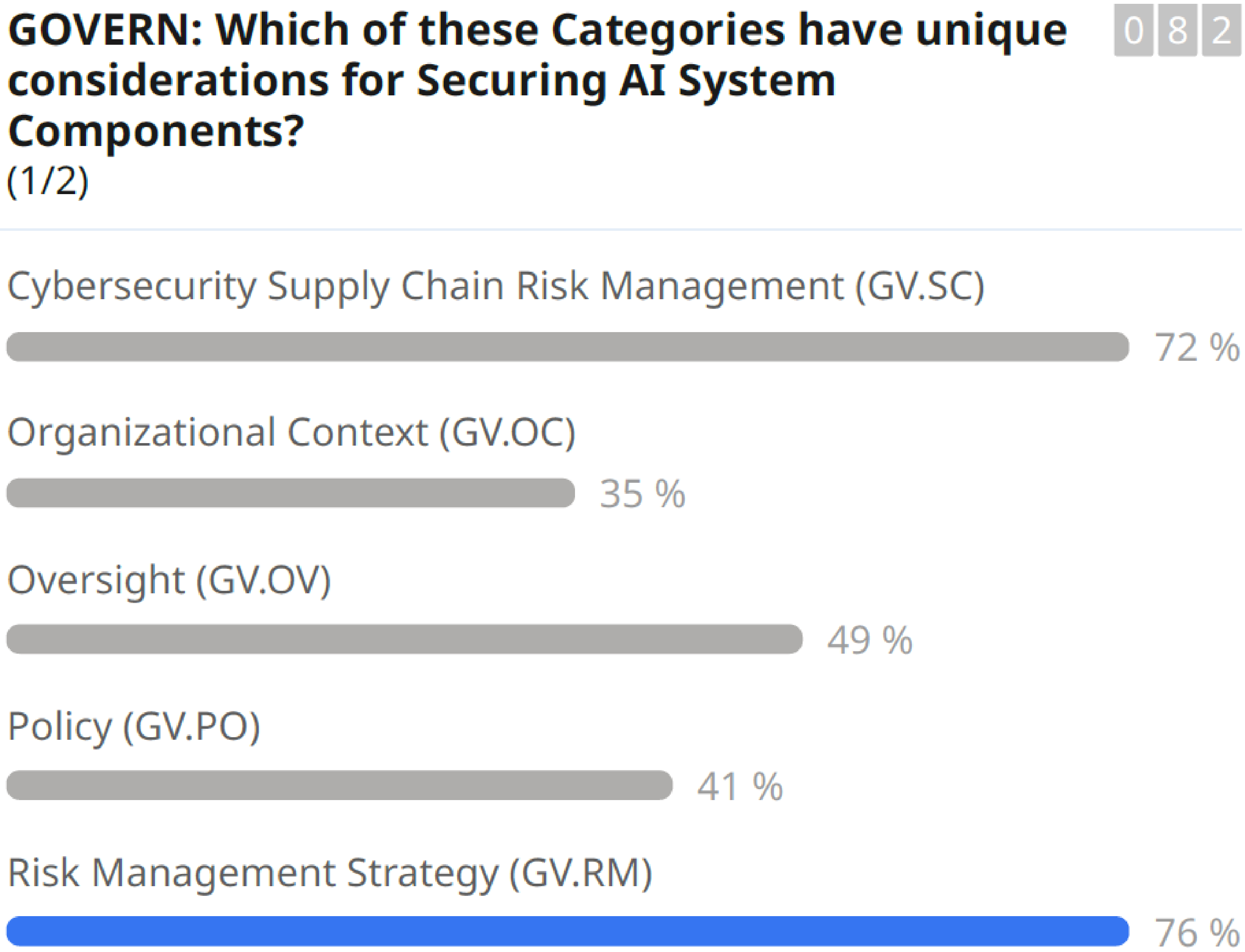
Govern

Category	Description	Heatmap
Cybersecurity Supply Chain Risk Management (GV.SC)	Cyber supply chain risk management processes are identified, established, managed, monitored, and improved by organizational stakeholders	1.0
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	0.40
Oversight (GV.OV)	Results of organization-wide cybersecurity risk management activities and performance are used to inform, improve, and adjust the risk management strategy	0.27
Policy (GV.PO)	Organizational cybersecurity policy is established, communicated, and enforced	0.12
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	0.56
Roles, Responsibilities, and Authorities (GV.RR)	Cybersecurity roles, responsibilities, and authorities to foster accountability, performance assessment, and continuous improvement are established and communicated	0.06

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of current emphasis): 🟥 **Low** 🟨 **Moderate** 🟩 **High**

Slido Results: *Govern (1 of 4)*



GOVERN: Which of these Categories have unique considerations for Securing AI System Components?
(1/2)

0 8 2

Cybersecurity Supply Chain Risk Management (GV.SC) 72 %

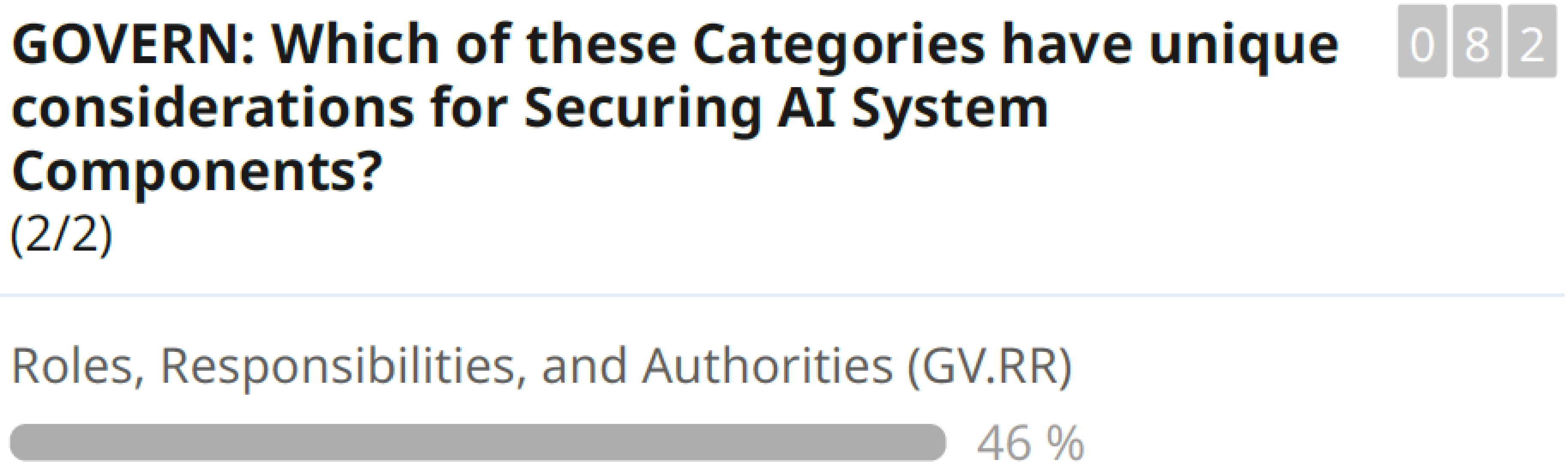
Organizational Context (GV.OC) 35 %

Oversight (GV.OV) 49 %

Policy (GV.PO) 41 %

Risk Management Strategy (GV.RM) 76 %

Slido Results: *Govern (2 of 4)*



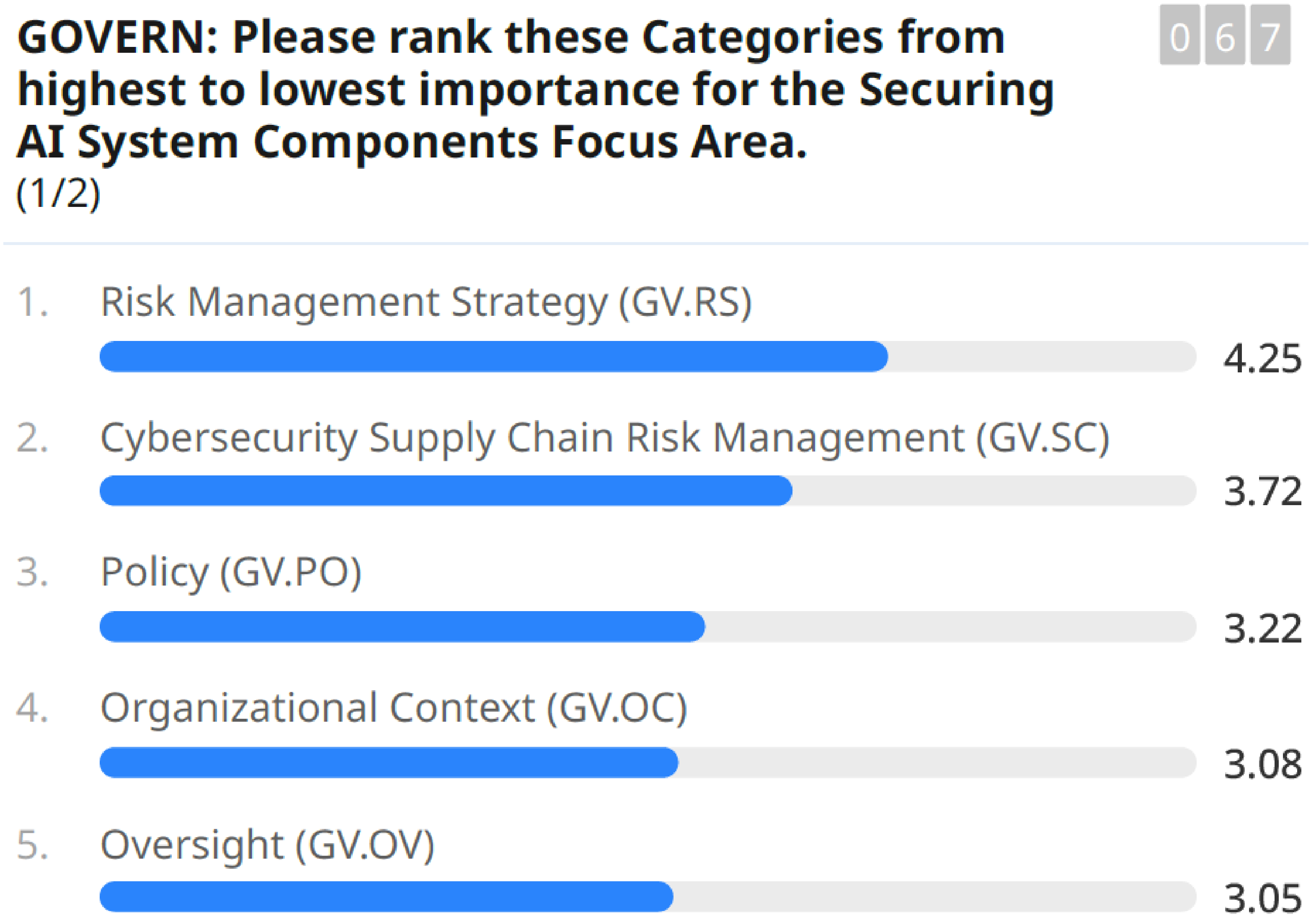
GOVERN: Which of these Categories have unique considerations for Securing AI System Components?
(2/2)

082

Roles, Responsibilities, and Authorities (GV.RR)

46 %

Slido Results: *Govern (3 of 4)*

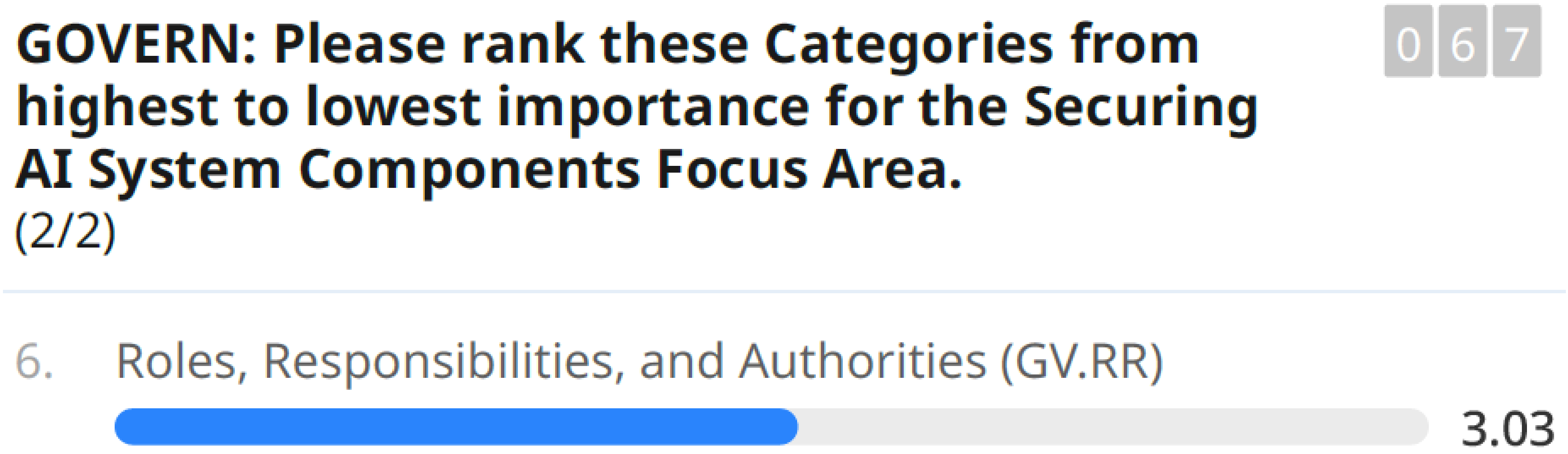


GOVERN: Please rank these Categories from highest to lowest importance for the Securing AI System Components Focus Area.
(1/2)

067

1. Risk Management Strategy (GV.RS) 4.25
2. Cybersecurity Supply Chain Risk Management (GV.SC) 3.72
3. Policy (GV.PO) 3.22
4. Organizational Context (GV.OC) 3.08
5. Oversight (GV.OV) 3.05

Slido Results: *Govern (4 of 4)*



GOVERN: Please rank these Categories from highest to lowest importance for the Securing AI System Components Focus Area.
(2/2)

6. Roles, Responsibilities, and Authorities (GV.RR) 3.03

Protect

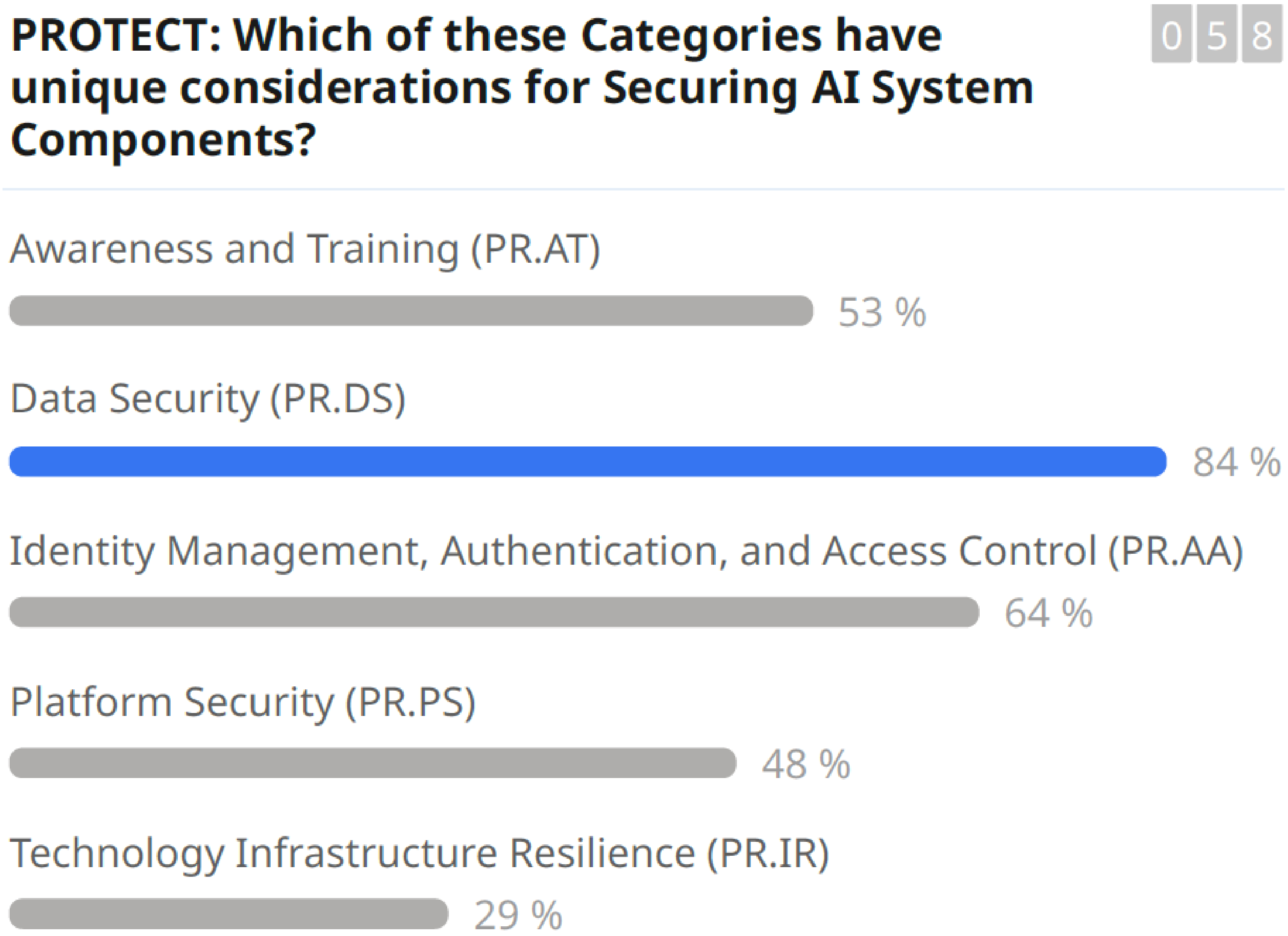
Category	Description	Heatmap
Awareness and Training (PR.AT)	The organization’s personnel are provided with cybersecurity awareness and training so that they can perform their cybersecurity-related tasks.	0.10
Data Security (PR.DS)	Data are managed consistent with the organization’s risk strategy to protect the confidentiality, integrity, and availability of information.	0.37
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.	0.42
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.	0.44
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.	0.31

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of current emphasis):

Low **Moderate** **High**

Slido Results: *Protect (1 of 2)*



PROTECT: Which of these Categories have unique considerations for Securing AI System Components?

Awareness and Training (PR.AT) 53 %

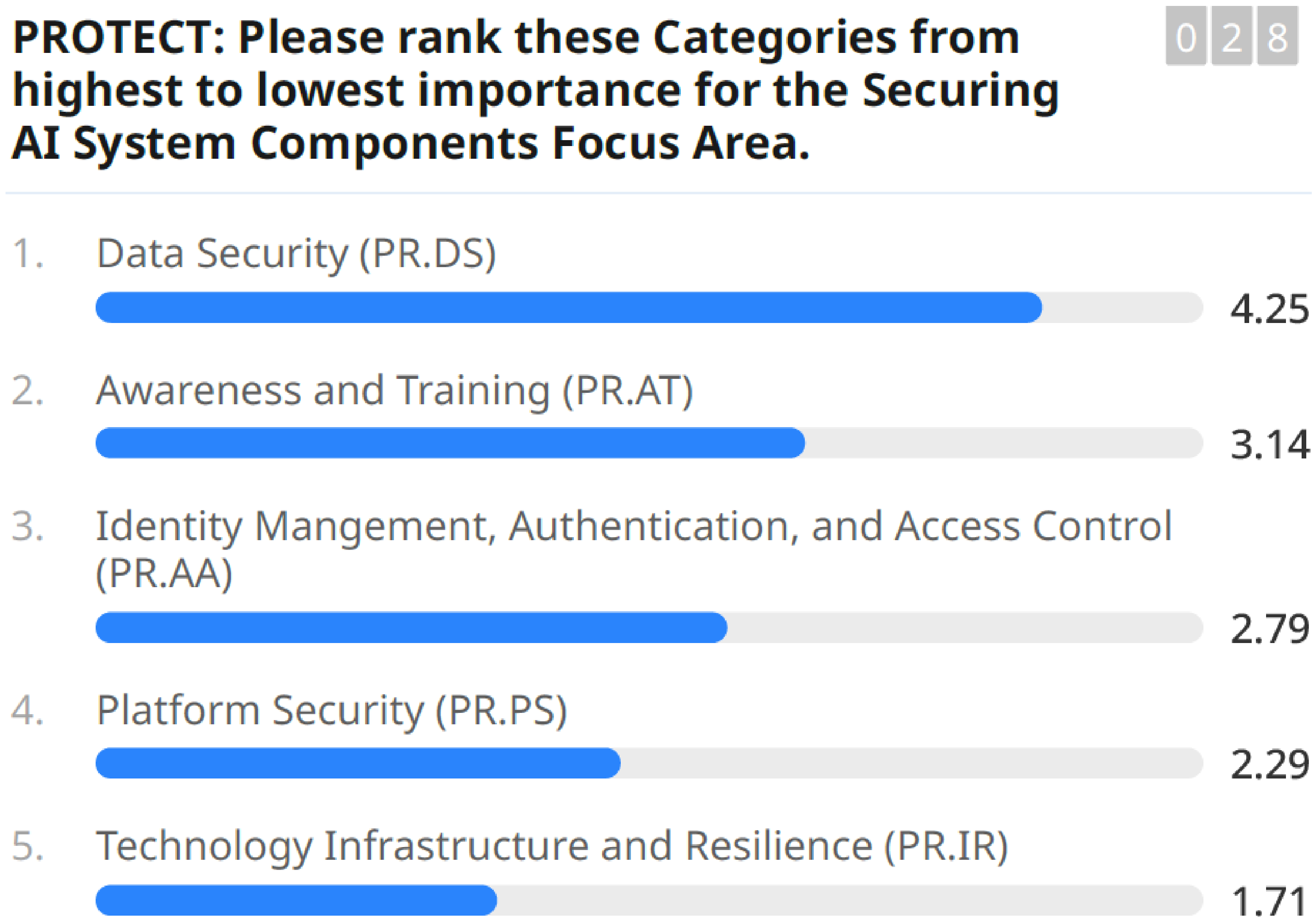
Data Security (PR.DS) 84 %

Identity Management, Authentication, and Access Control (PR.AA) 64 %

Platform Security (PR.PS) 48 %

Technology Infrastructure Resilience (PR.IR) 29 %

Slido Results: *Protect (2 of 2)*



PROTECT: Please rank these Categories from highest to lowest importance for the Securing AI System Components Focus Area.

028

1. Data Security (PR.DS) 4.25
2. Awareness and Training (PR.AT) 3.14
3. Identity Mangement, Authentication, and Access Control (PR.AA) 2.79
4. Platform Security (PR.PS) 2.29
5. Technology Infrastructure and Resilience (PR.IR) 1.71

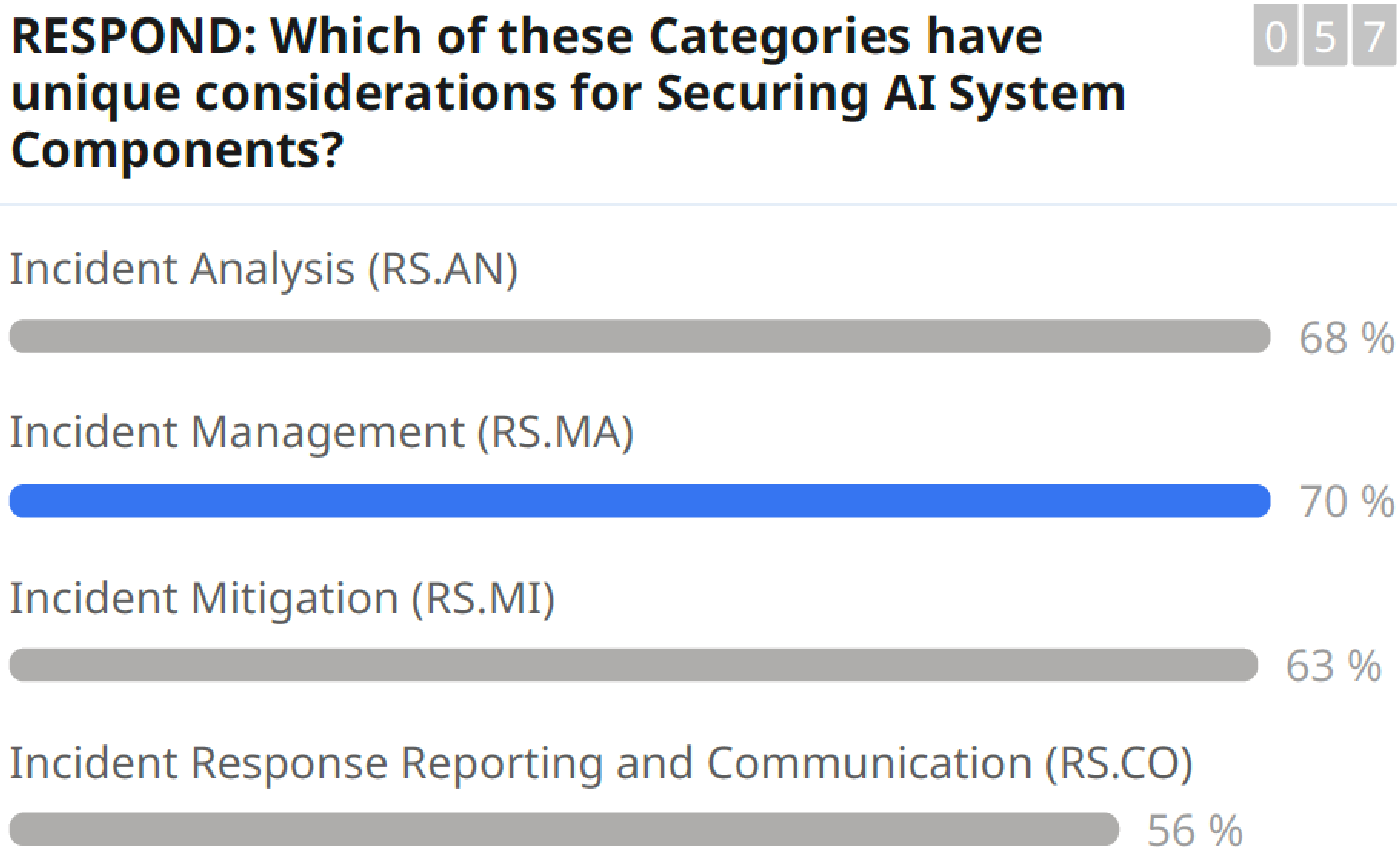
Respond

Category	Description	Heatmap
Incident Analysis (RS.AN)	Investigations are conducted to ensure effective response and support forensics and recovery activities.	0.15
Incident Management (RS.MA)	Responses to detected cybersecurity incidents are managed.	0.19
Incident Mitigation (RS.MI)	Activities are performed to prevent expansion of an event and mitigate its effects.	0.02
Incident Response Reporting and Communication (RS.CO)	Response activities are coordinated with internal and external stakeholders as required by laws, regulations, or policies.	0.17

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of current emphasis): **Low** **Moderate** **High**

Slido Results: *Respond (1 of 2)*



RESPONDS: Which of these Categories have unique considerations for Securing AI System Components?

057

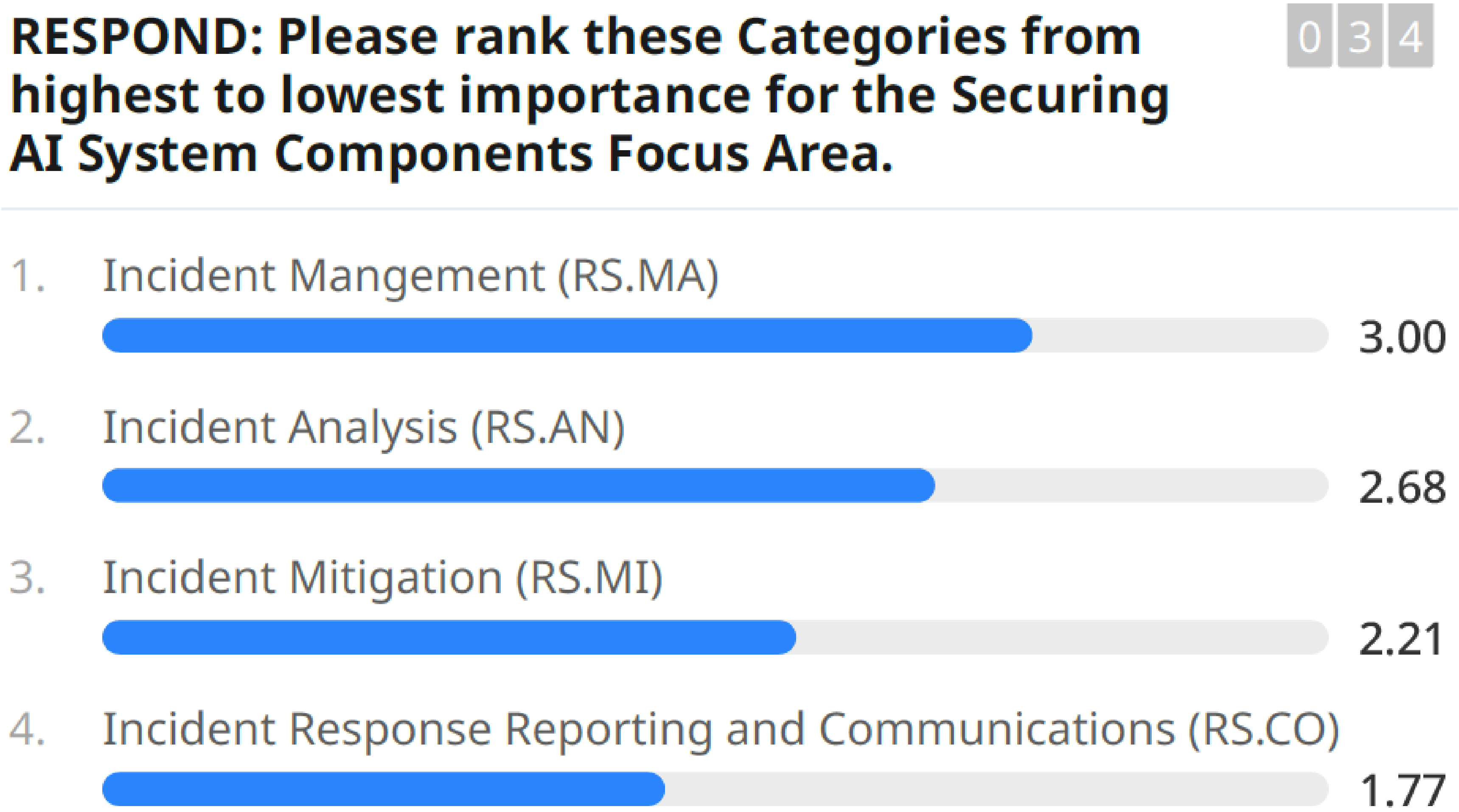
Incident Analysis (RS.AN) 68 %

Incident Management (RS.MA) 70 %

Incident Mitigation (RS.MI) 63 %

Incident Response Reporting and Communication (RS.CO) 56 %

Slido Results: *Respond (2 of 2)*



RESPOND: Please rank these Categories from highest to lowest importance for the Securing AI System Components Focus Area.

1. Incident Mangement (RS.MA) 3.00
2. Incident Analysis (RS.AN) 2.68
3. Incident Mitigation (RS.MI) 2.21
4. Incident Response Reporting and Communications (RS.CO) 1.77

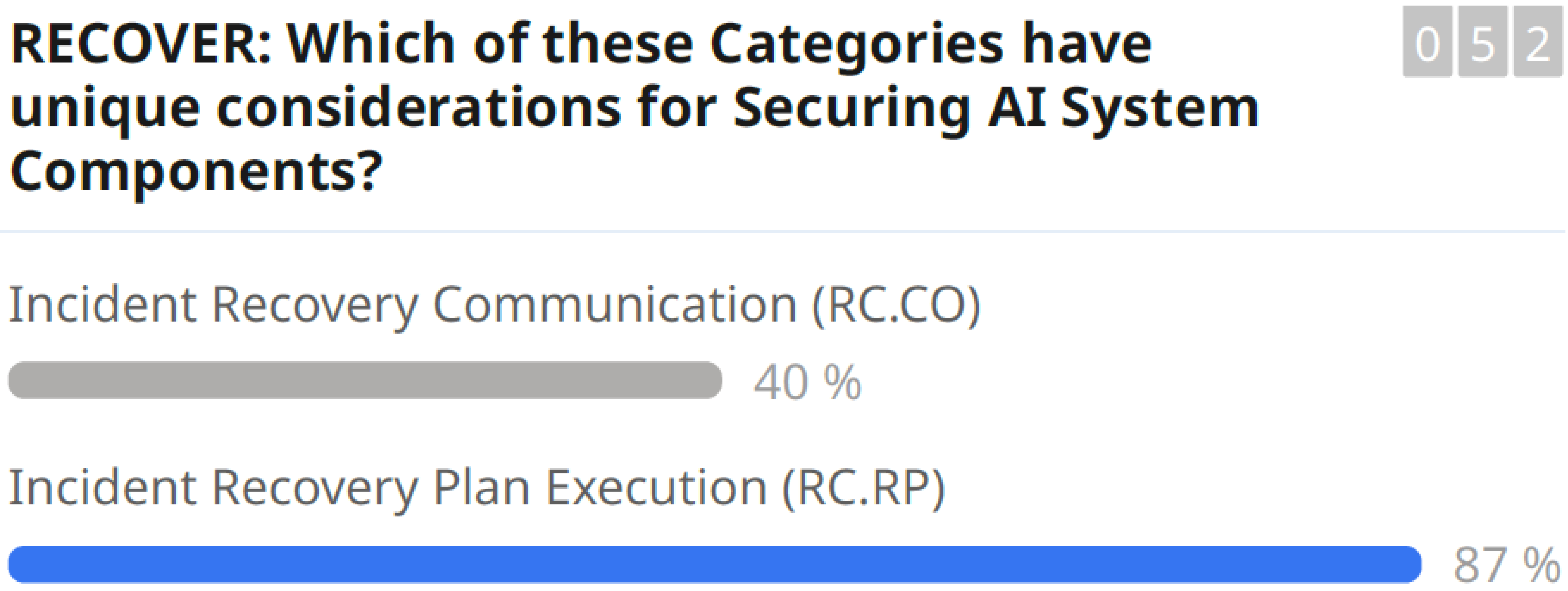
Recover

Category	Description	Heatmap
Incident Recovery Communications (RC.CO)	Restoration activities are coordinated with internal and external parties.	0.10
Incident Recovery Plan Execution (RC.RP)	Restoration activities are performed to ensure operational availability of systems and services affected by cybersecurity incidents.	0.13

FOR DISCUSSION PURPOSES ONLY

Heatmap Legend 0-1 (degree of current emphasis): 🟥 **Low** 🟨 **Moderate** 🟩 **High**

Slido Results: *Recover (1 of 2)*

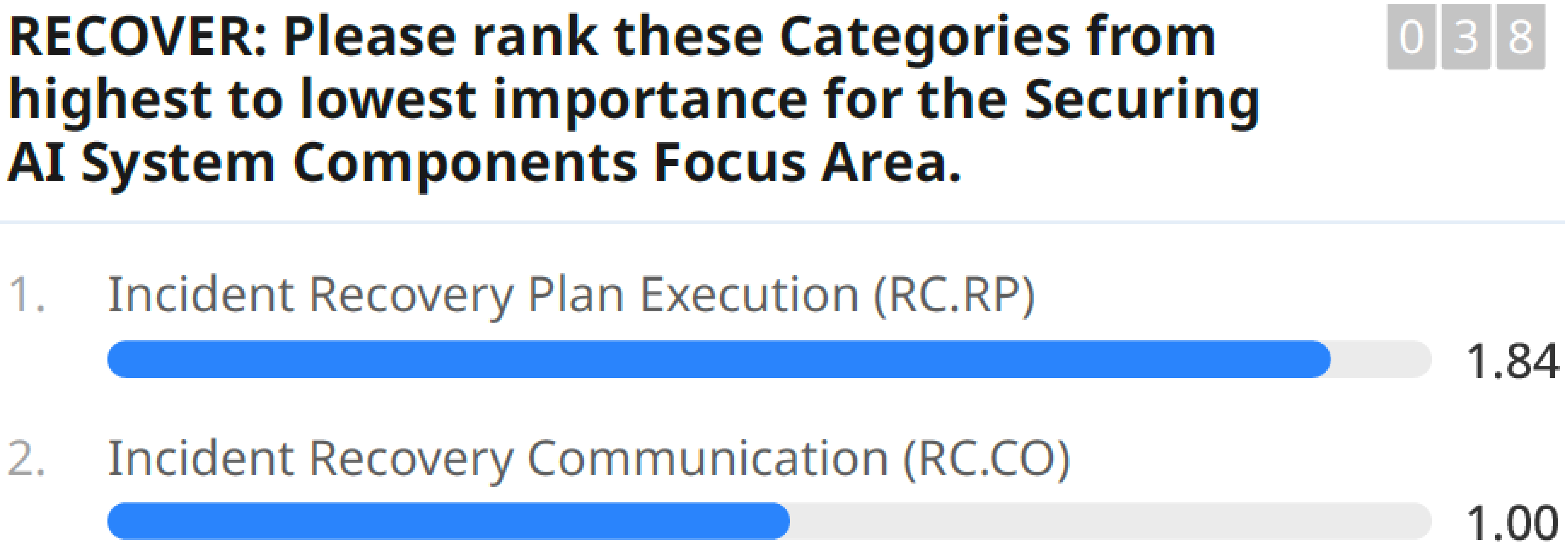


RECOVER: Which of these Categories have unique considerations for Securing AI System Components?

Incident Recovery Communication (RC.CO) 40 %

Incident Recovery Plan Execution (RC.RP) 87 %

Slido Results: *Recover (2 of 2)*



RECOVER: Please rank these Categories from highest to lowest importance for the Securing AI System Components Focus Area.

1. Incident Recovery Plan Execution (RC.RP) 1.84
2. Incident Recovery Communication (RC.CO) 1.00

Describing the Securing AI System Components Focus Area

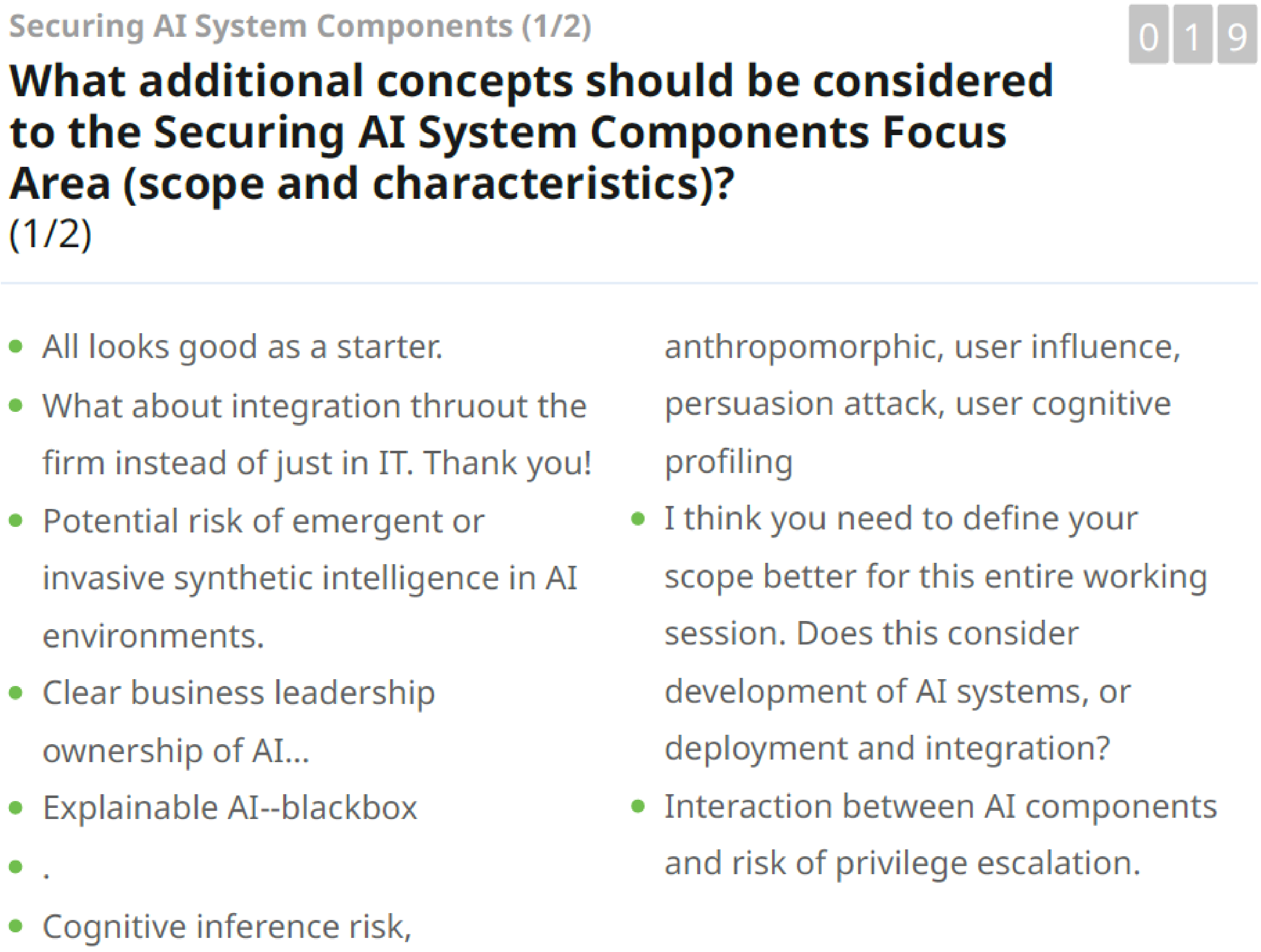
- **Scope:** Identifying cybersecurity challenges when integrating AI into organizational ecosystems and infrastructure
- **Focus Area characteristics – cybersecurity considerations regarding:**
  - Preparing an organization to develop and/or deploy AI capabilities
  - Managing and AI system throughout is lifecycle
  - Understanding the threat and vulnerability landscape for AI technology components
- **Examples of AI Technology System Components:**
  - AI Model (e.g., LLMs, computer vision models, RL)
  - Data pipeline and storage (e.g., understanding provenance, training, cleaning, anomalies)
  - System Interfaces, internal tool callings (e.g., AI agents, interfaces)
  - Third-party AI platforms/services, external tool callings (e.g., LLMs, AI agents, APIs)

Challenge Areas

- Data Governance, Security, and Privacy
- Adversarial Attacks
- Unauthorized Access and Use
- Identifying, Detecting, and Responding to AI Component Vulnerabilities and Adverse Events
- Supply Chain Security
- Model Drift and Onset of Unexpected or Inaccurate Results

Slido.com
#CyberAI_WS1

Slido Results: *Securing AI System Components Focus Area (1 of 7)*



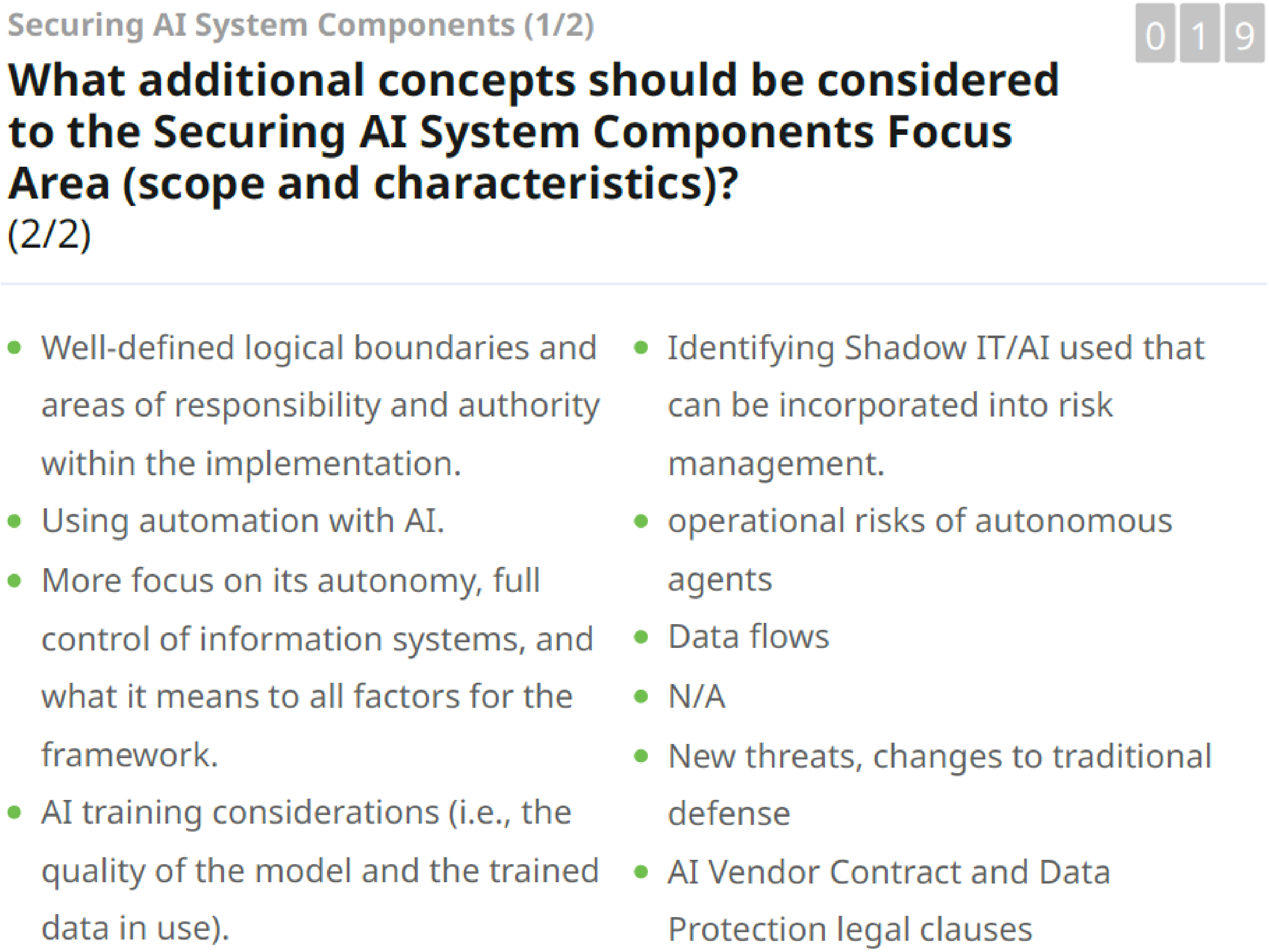
Securing AI System Components (1/2)

019

What additional concepts should be considered to the Securing AI System Components Focus Area (scope and characteristics)?
(1/2)

- All looks good as a starter.
- What about integration thruout the firm instead of just in IT. Thank you!
- Potential risk of emergent or invasive synthetic intelligence in AI environments.
- Clear business leadership ownership of AI...
- Explainable AI--blackbox
- .
- Cognitive inference risk, anthropomorphic, user influence, persuasion attack, user cognitive profiling
- I think you need to define your scope better for this entire working session. Does this consider development of AI systems, or deployment and integration?
- Interaction between AI components and risk of privilege escalation.

Slido Results: *Securing AI System Components Focus Area (2 of 7)*



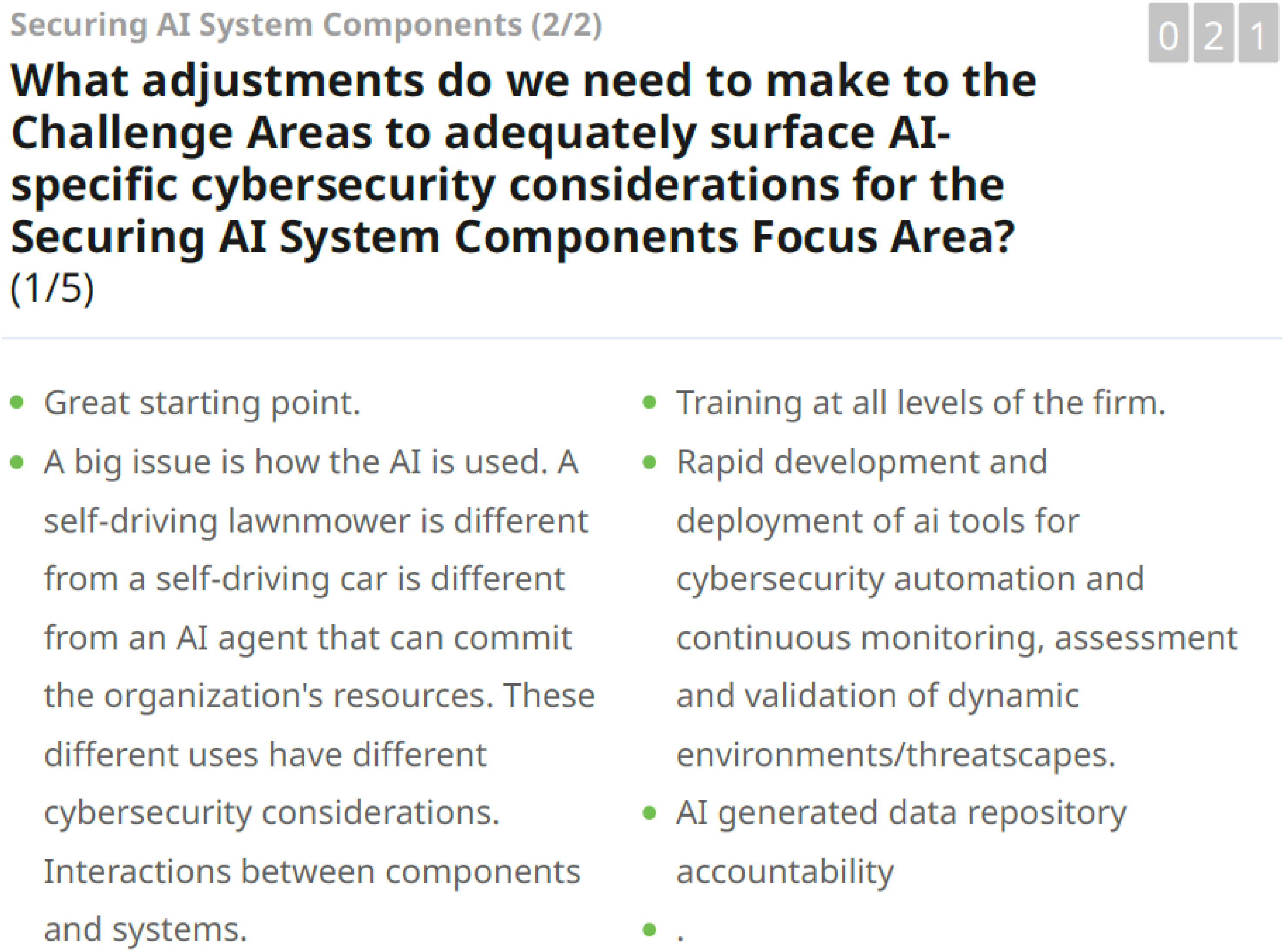
Securing AI System Components (1/2)

019

What additional concepts should be considered to the Securing AI System Components Focus Area (scope and characteristics)?
(2/2)

- Well-defined logical boundaries and areas of responsibility and authority within the implementation.
- Using automation with AI.
- More focus on its autonomy, full control of information systems, and what it means to all factors for the framework.
- AI training considerations (i.e., the quality of the model and the trained data in use).
- Identifying Shadow IT/AI used that can be incorporated into risk management.
- operational risks of autonomous agents
- Data flows
- N/A
- New threats, changes to traditional defense
- AI Vendor Contract and Data Protection legal clauses

Slido Results: *Securing AI System Components Focus Area (3 of 7)*



Securing AI System Components (2/2)

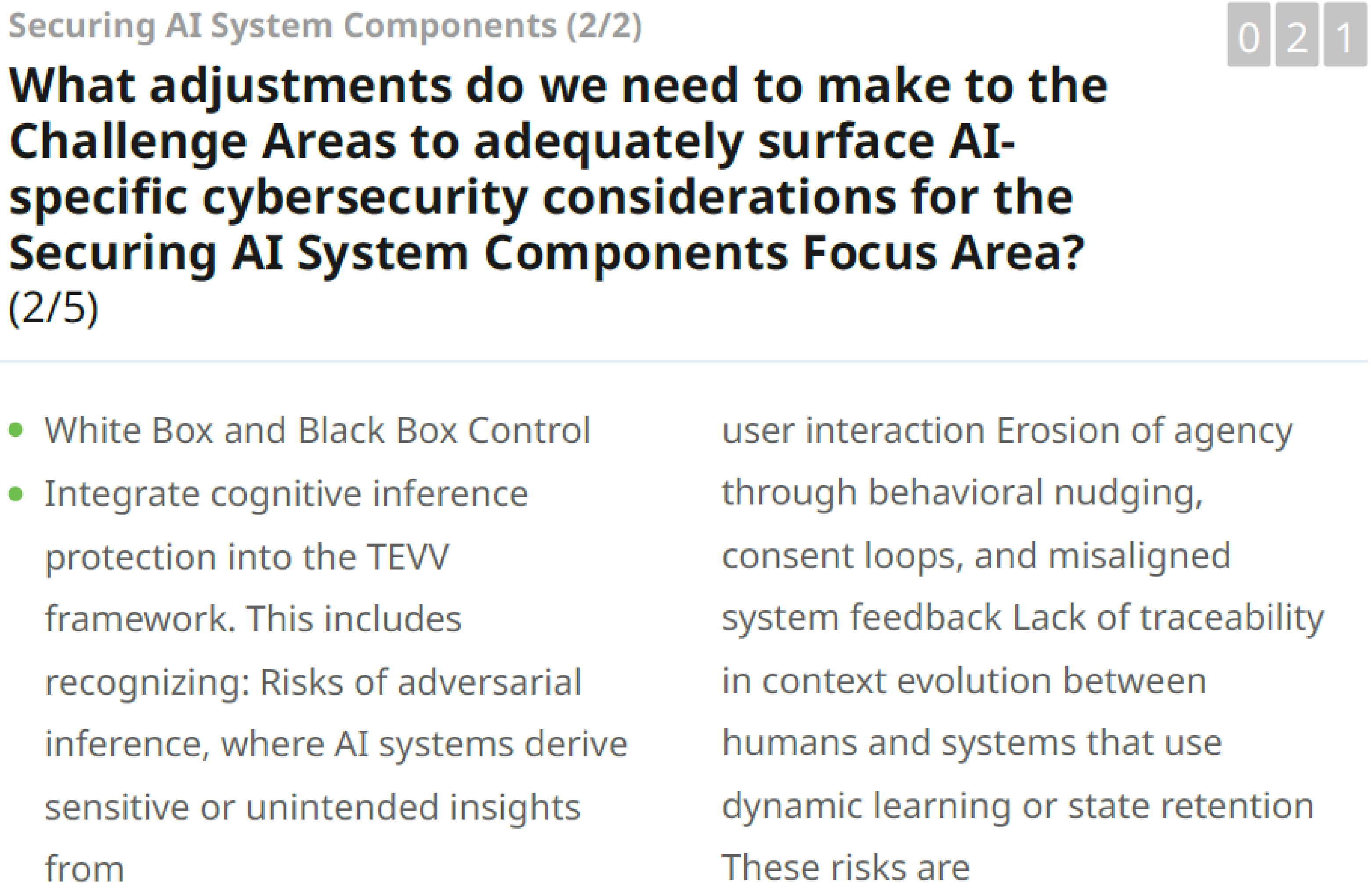
021

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations for the Securing AI System Components Focus Area?

(1/5)

- Great starting point.
- A big issue is how the AI is used. A self-driving lawnmower is different from a self-driving car is different from an AI agent that can commit the organization's resources. These different uses have different cybersecurity considerations. Interactions between components and systems.
- Training at all levels of the firm.
- Rapid development and deployment of ai tools for cybersecurity automation and continuous monitoring, assessment and validation of dynamic environments/threatscapes.
- AI generated data repository accountability
- .

Slido Results: *Securing AI System Components Focus Area (4 of 7)*



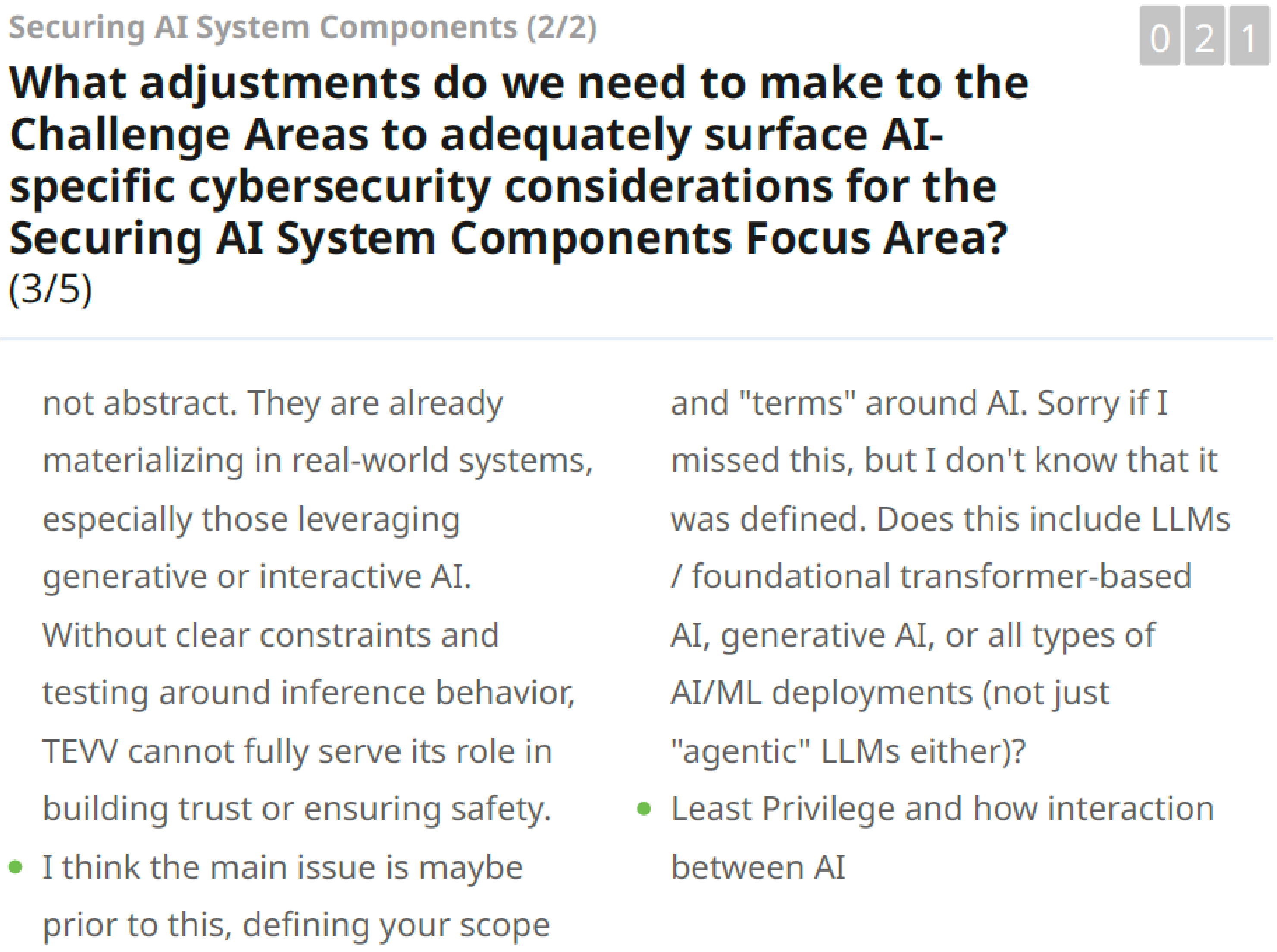
Securing AI System Components (2/2)

0 2 1

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations for the Securing AI System Components Focus Area?
(2/5)

- White Box and Black Box Control
- Integrate cognitive inference protection into the TEVV framework. This includes recognizing: Risks of adversarial inference, where AI systems derive sensitive or unintended insights from user interaction Erosion of agency through behavioral nudging, consent loops, and misaligned system feedback Lack of traceability in context evolution between humans and systems that use dynamic learning or state retention These risks are

Slido Results: *Securing AI System Components Focus Area (5 of 7)*



Securing AI System Components (2/2)

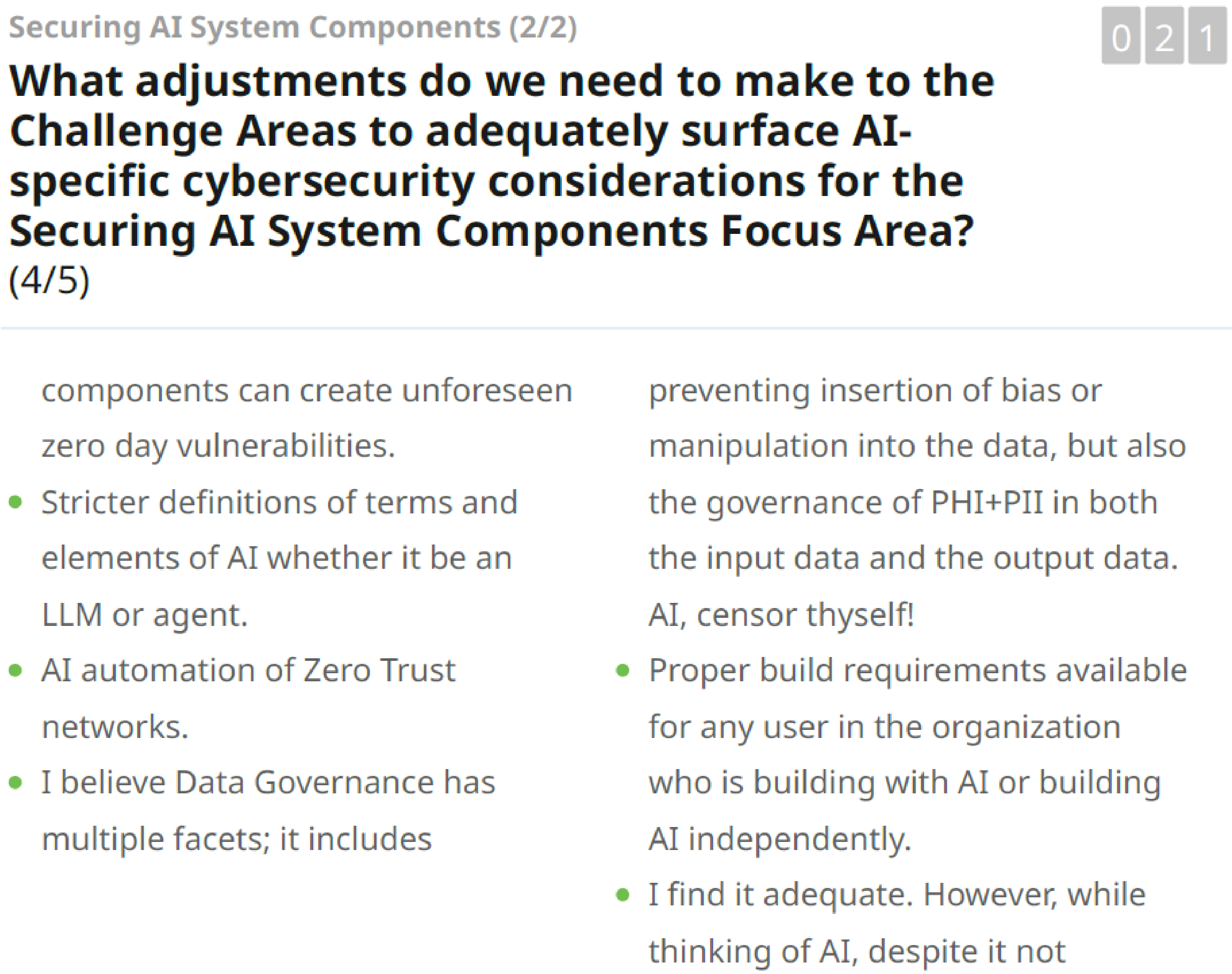
What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations for the Securing AI System Components Focus Area?

(3/5)

not abstract. They are already materializing in real-world systems, especially those leveraging generative or interactive AI. Without clear constraints and testing around inference behavior, TEVV cannot fully serve its role in building trust or ensuring safety.

- I think the main issue is maybe prior to this, defining your scope and "terms" around AI. Sorry if I missed this, but I don't know that it was defined. Does this include LLMs / foundational transformer-based AI, generative AI, or all types of AI/ML deployments (not just "agentic" LLMs either)?
- Least Privilege and how interaction between AI

Slido Results: *Securing AI System Components Focus Area (6 of 7)*



Securing AI System Components (2/2)

021

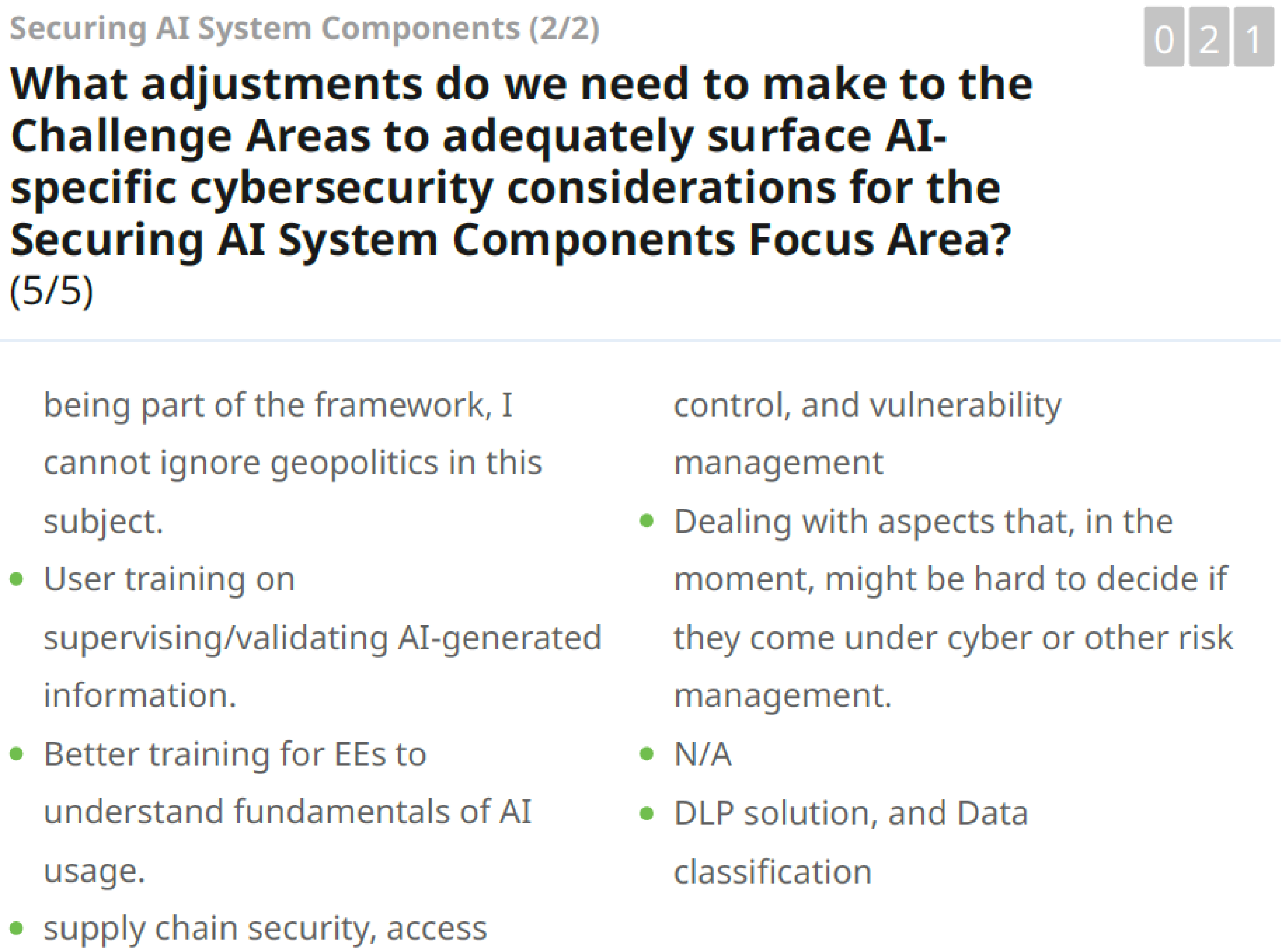
What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations for the Securing AI System Components Focus Area?

(4/5)

components can create unforeseen zero day vulnerabilities.

- Stricter definitions of terms and elements of AI whether it be an LLM or agent.
- AI automation of Zero Trust networks.
- I believe Data Governance has multiple facets; it includes preventing insertion of bias or manipulation into the data, but also the governance of PHI+PII in both the input data and the output data. AI, censor thyself!
- Proper build requirements available for any user in the organization who is building with AI or building AI independently.
- I find it adequate. However, while thinking of AI, despite it not

Slido Results: *Securing AI System Components Focus Area (7 of 7)*



Securing AI System Components (2/2)

021

What adjustments do we need to make to the Challenge Areas to adequately surface AI-specific cybersecurity considerations for the Securing AI System Components Focus Area?

(5/5)

being part of the framework, I cannot ignore geopolitics in this subject.

- User training on supervising/validating AI-generated information.
- Better training for EEs to understand fundamentals of AI usage.
- supply chain security, access control, and vulnerability management
- Dealing with aspects that, in the moment, might be hard to decide if they come under cyber or other risk management.
- N/A
- DLP solution, and Data classification

Close-out

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

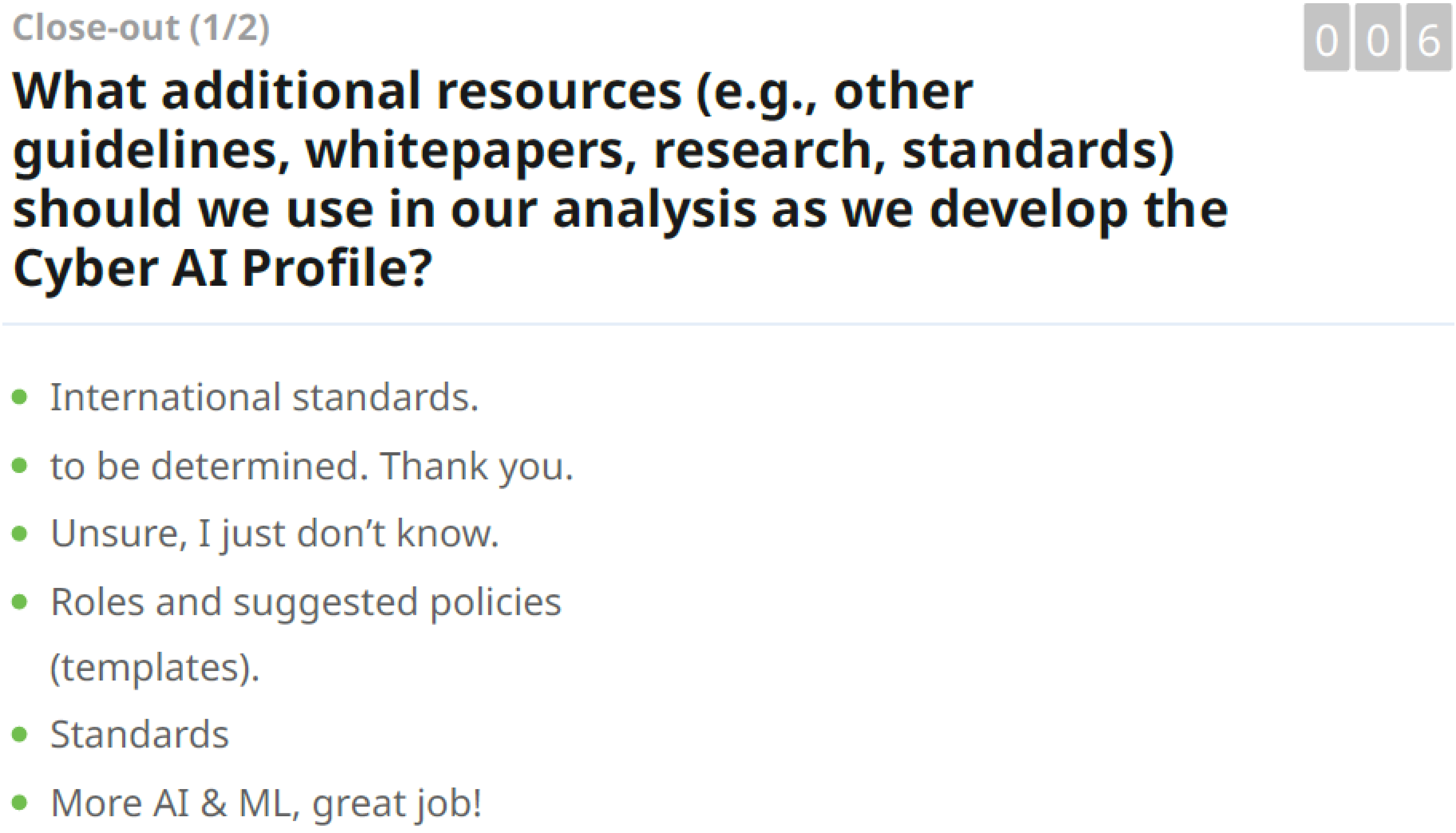
We Appreciate Your Input

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_WS1

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



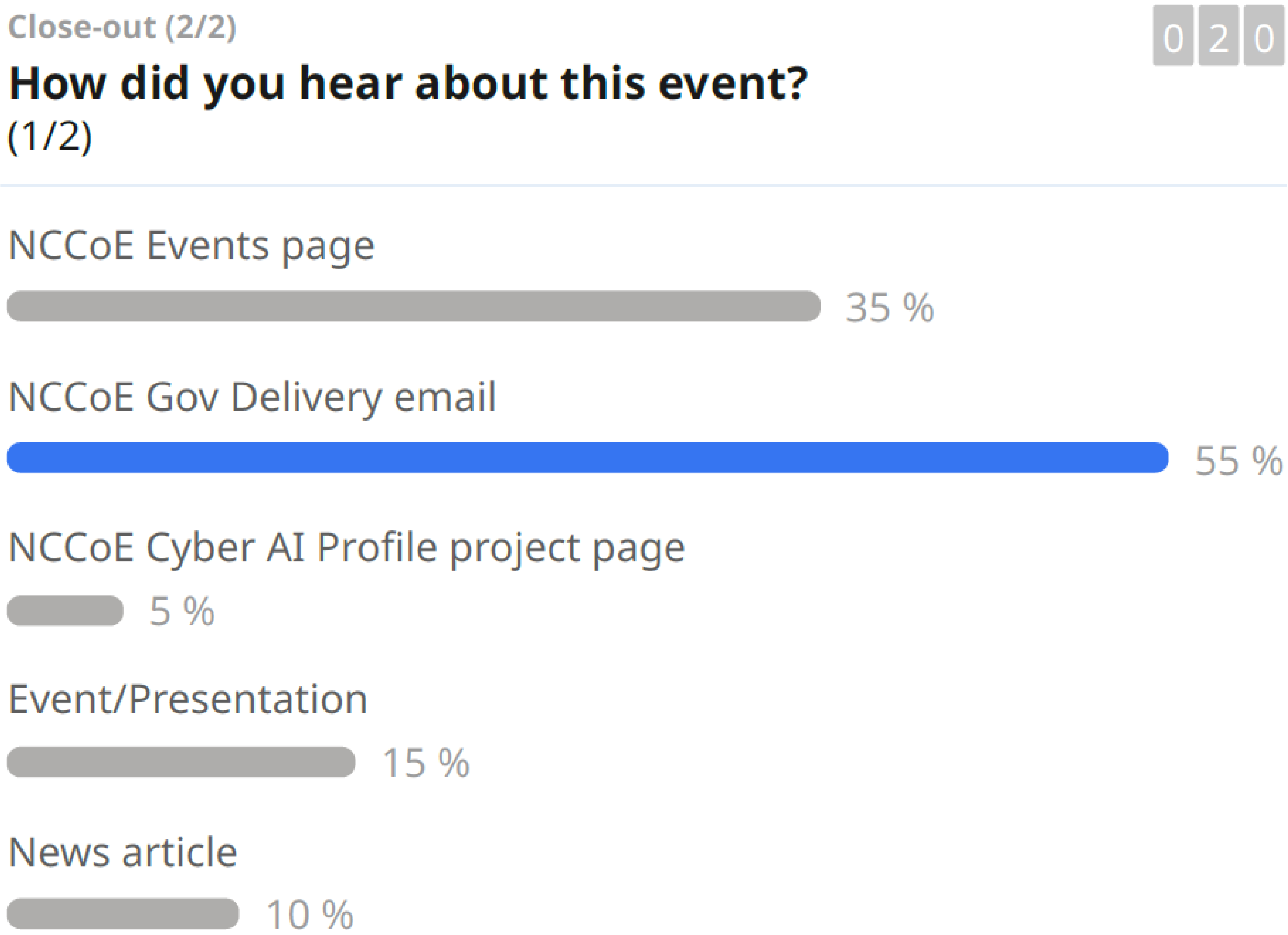
Close-out (1/2)

006

What additional resources (e.g., other guidelines, whitepapers, research, standards) should we use in our analysis as we develop the Cyber AI Profile?

- International standards.
- to be determined. Thank you.
- Unsure, I just don't know.
- Roles and suggested policies (templates).
- Standards
- More AI & ML, great job!

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



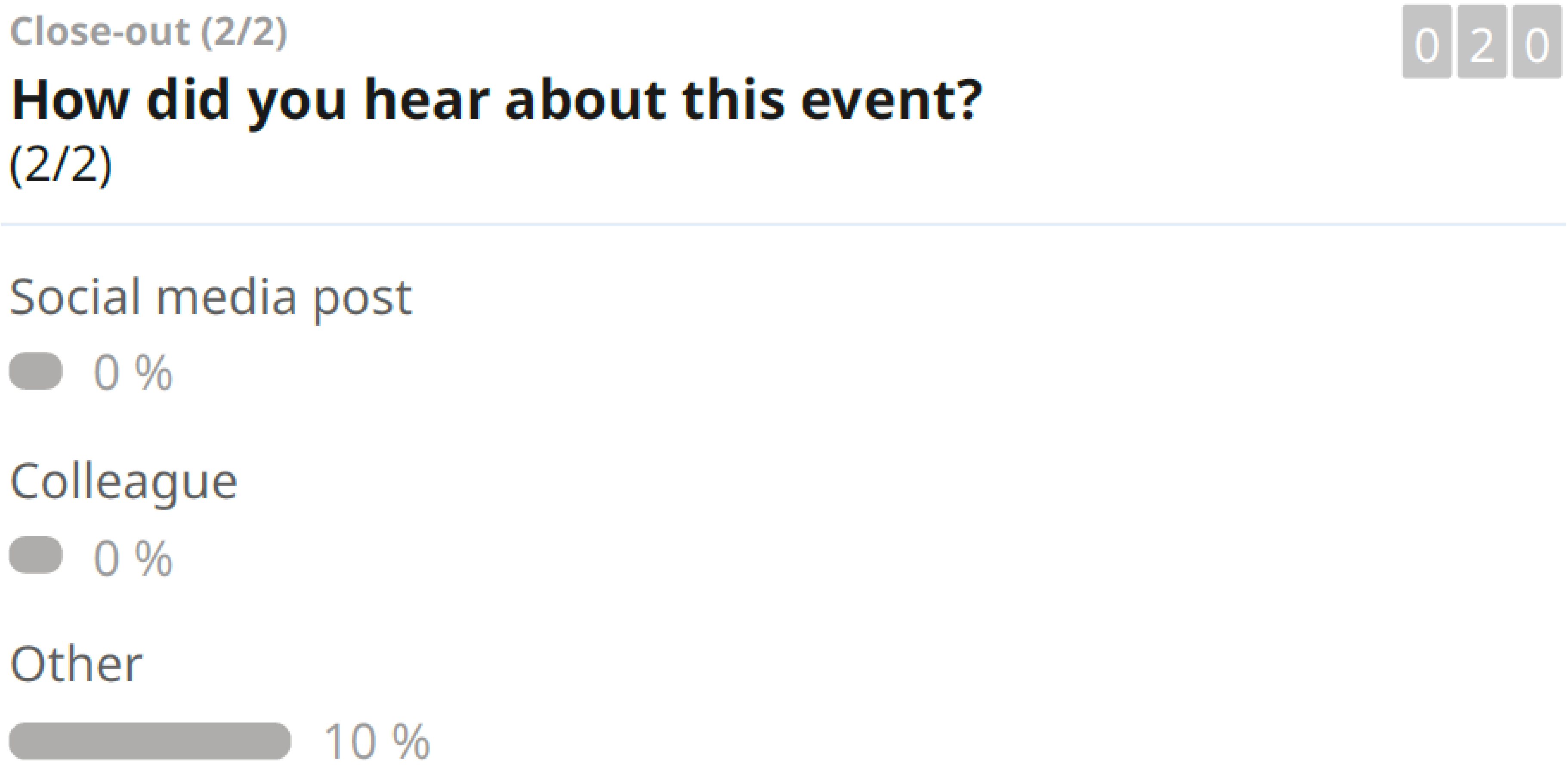
Close-out (2/2)

020

How did you hear about this event?
(1/2)

NCCoE Events page	35 %
NCCoE Gov Delivery email	55 %
NCCoE Cyber AI Profile project page	5 %
Event/Presentation	15 %
News article	10 %

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



Close-out (2/2)

0 2 0

How did you hear about this event?
(2/2)

Social media post

0 %

Colleague

0 %

Other

10 %

Working Session Schedule

August 19, 2025

Conducting AI-enabled Cyber Defense

☑

August 26, 2025

Securing AI System Components

☑

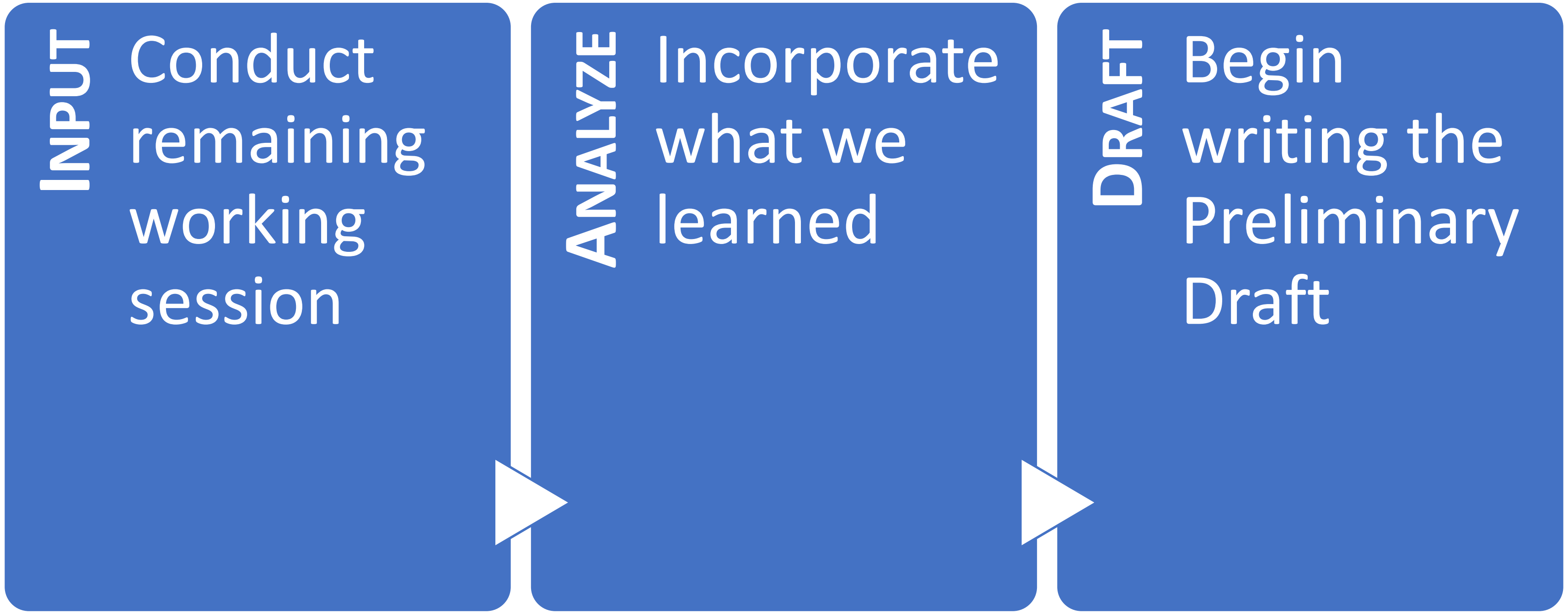
September 2, 2025

1:00 – 4:00pm EDT

Thwarting AI-enabled Cyber Attacks

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Working Sessions Next Steps



INPUT

Conduct remaining working session

ANALYZE

Incorporate what we learned

DRAFT

Begin writing the Preliminary Draft

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!

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