Cyber AI Profile Working Sessions: Thwarting AI-enabled Cyber Attacks

September 2, 2025

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

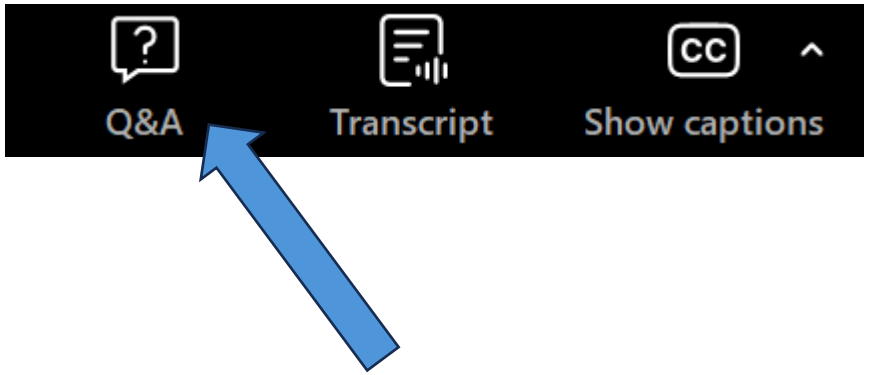
Agenda

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

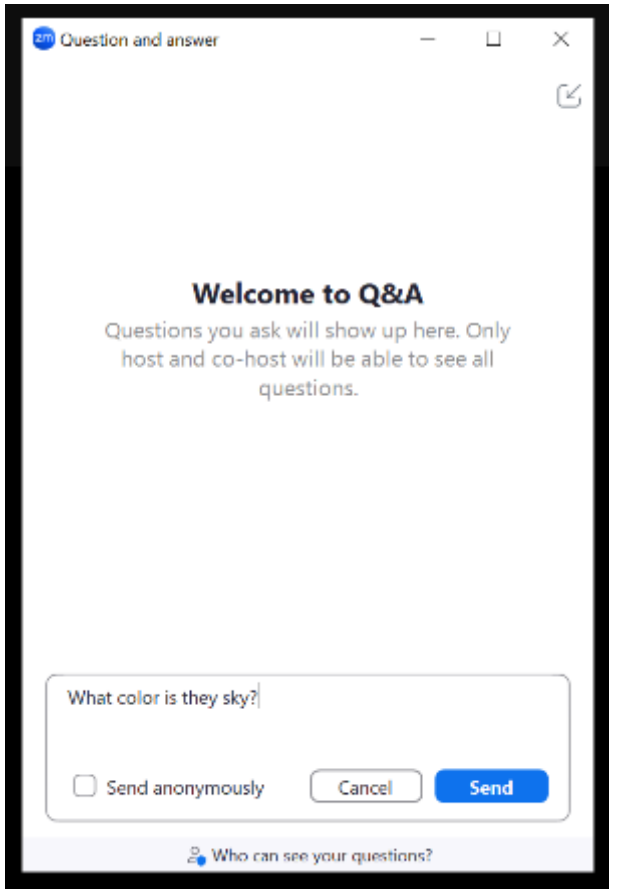
Submitting Questions

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

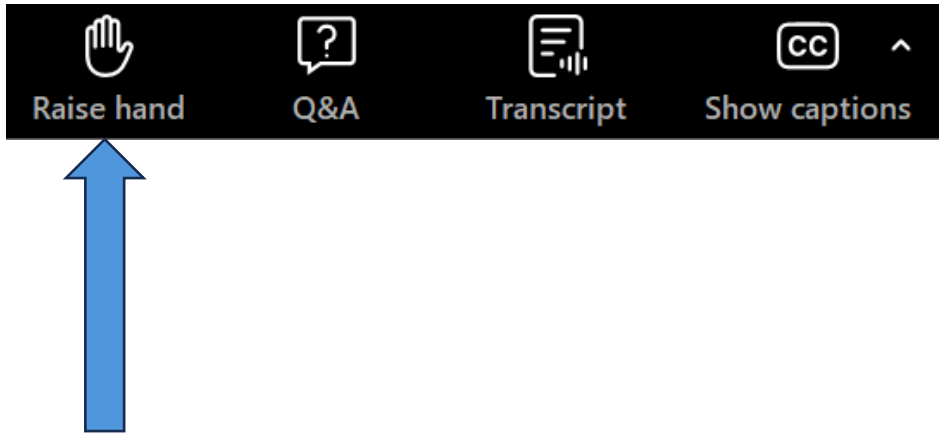


2. Type your question in the text box and click Send

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 (Nov 6 – Tentative) (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

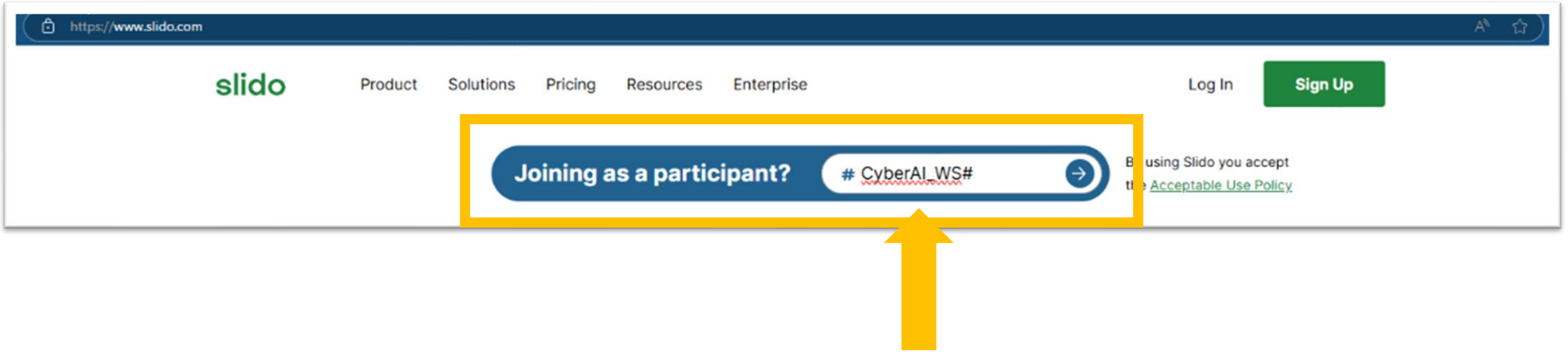
NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

How You Contribute Today

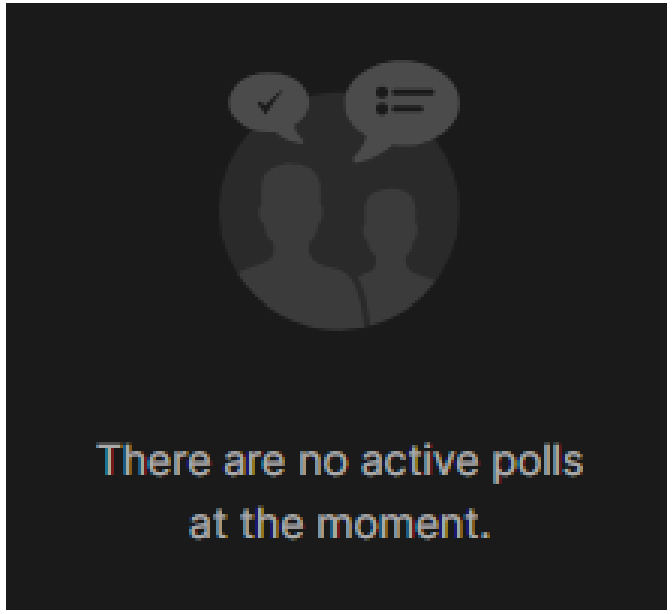
- **Please raise your virtual hand or type in the chat to contribute**
- Members of the press, please identify yourself and your organization
- Be respectful of others
- Please don’t be shy – we would love to hear from everyone!
- Please remain on mute when not speaking
- We will use Slido to facilitate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

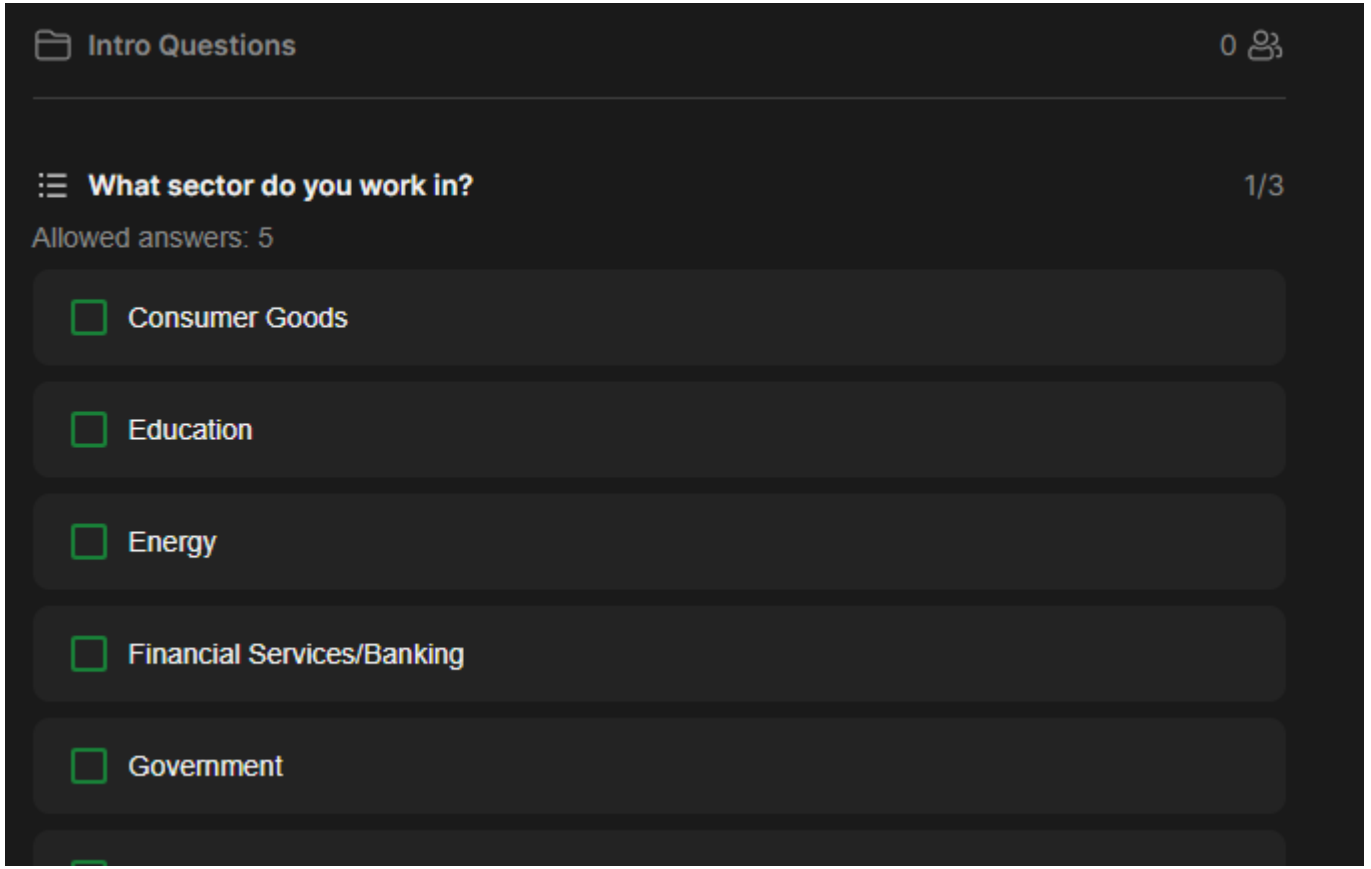
Using Slido

- 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



Slido Screens



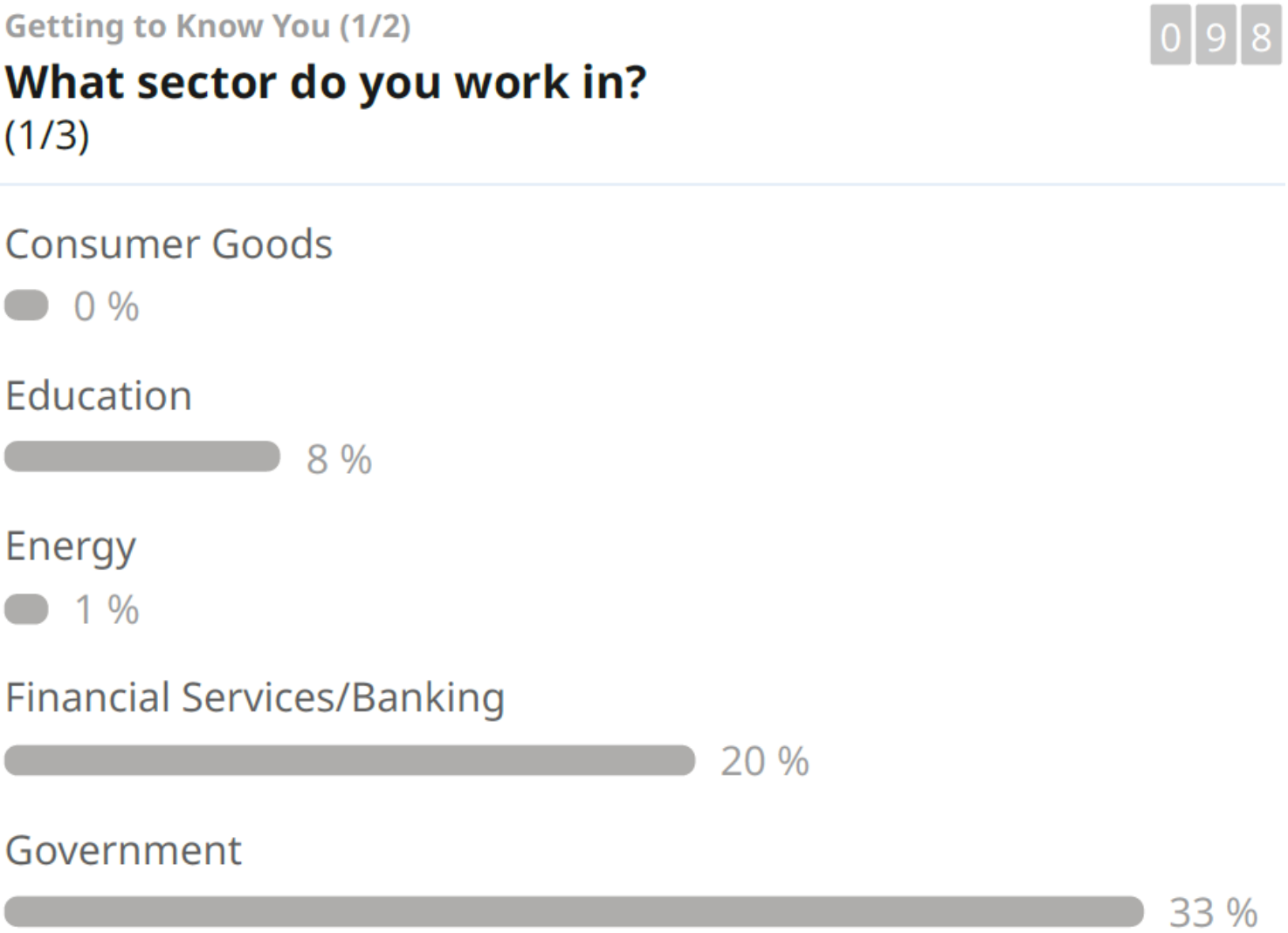


Slido: Getting to Know You

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

Slido.com
#CyberAI_WS3

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)*

Getting to Know You (1/2)

0 9 8

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

Telecommunications
2 %

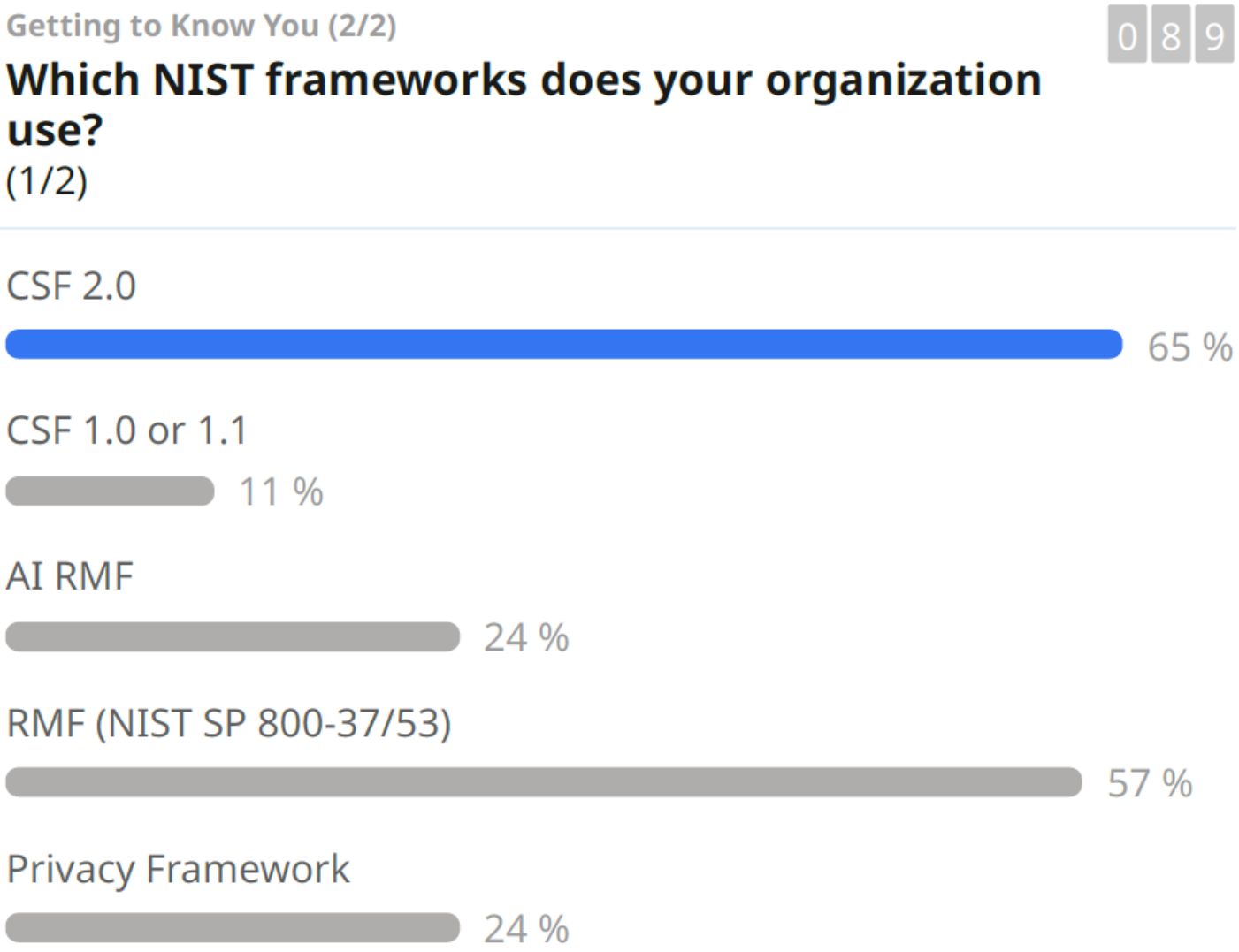
Think Tank
2 %

Trade Association
1 %

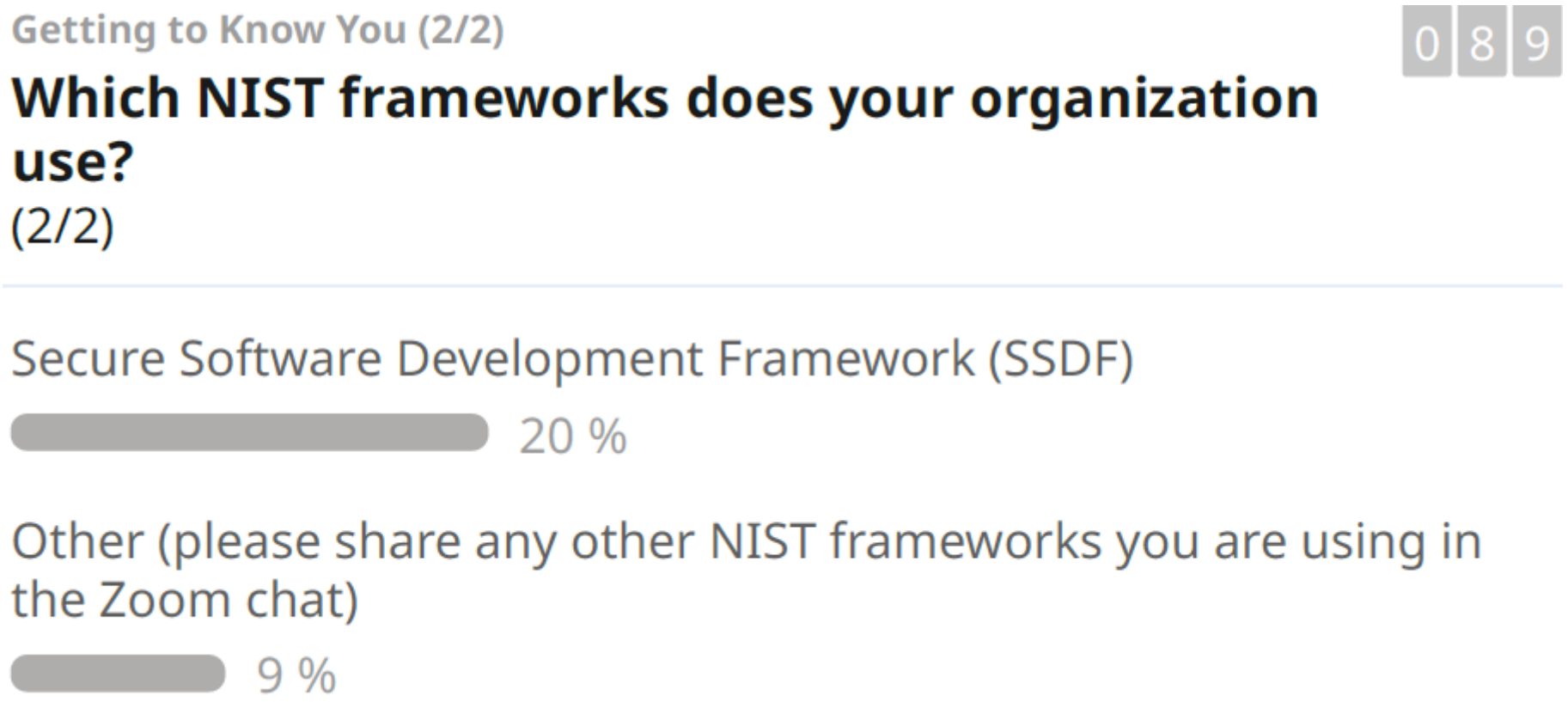
Transportation
1 %

Other (please share your sector in the Zoom chat)
2 %

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 Thwarting AI-Enabled Cyber Attacks

- **Scope:** Bolstering defenses and building resilience to protect against new AI-enabled threat vectors.
- **Focus Area characteristics – cybersecurity considerations regarding:**
  - Understand adversarial uses of AI
  - Preparing an organization to defend against AI-enabled cyber attacks
  - Identifying and managing AI-related threats
- **Examples of AI-enabled Cyber Attacks:**
  - Self-learning/adaptive malware
  - Automated vulnerability scanning
  - Optimizing botnet coordination
  - Automating IoT activities for exploitation

Challenge Areas

- Automated and Adaptive Malware
- Phishing and Social Engineering
- Credential Stuffing and Brute Force Attacks
- AI-Driven Reconnaissance and Cyber Espionage
- Distributed Denial of Service (DDoS) Attacks
- Evasion Techniques
- Supply Chain Attacks
- AI-Powered Zero-Day Exploits
- AI-Augmented Physical Cyber Attacks
- AI-Powered Attack Automation

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 cybersecurity capabilities do we need to enhance to better position ourselves to thwart AI-enabled cyber attacks?
    - 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

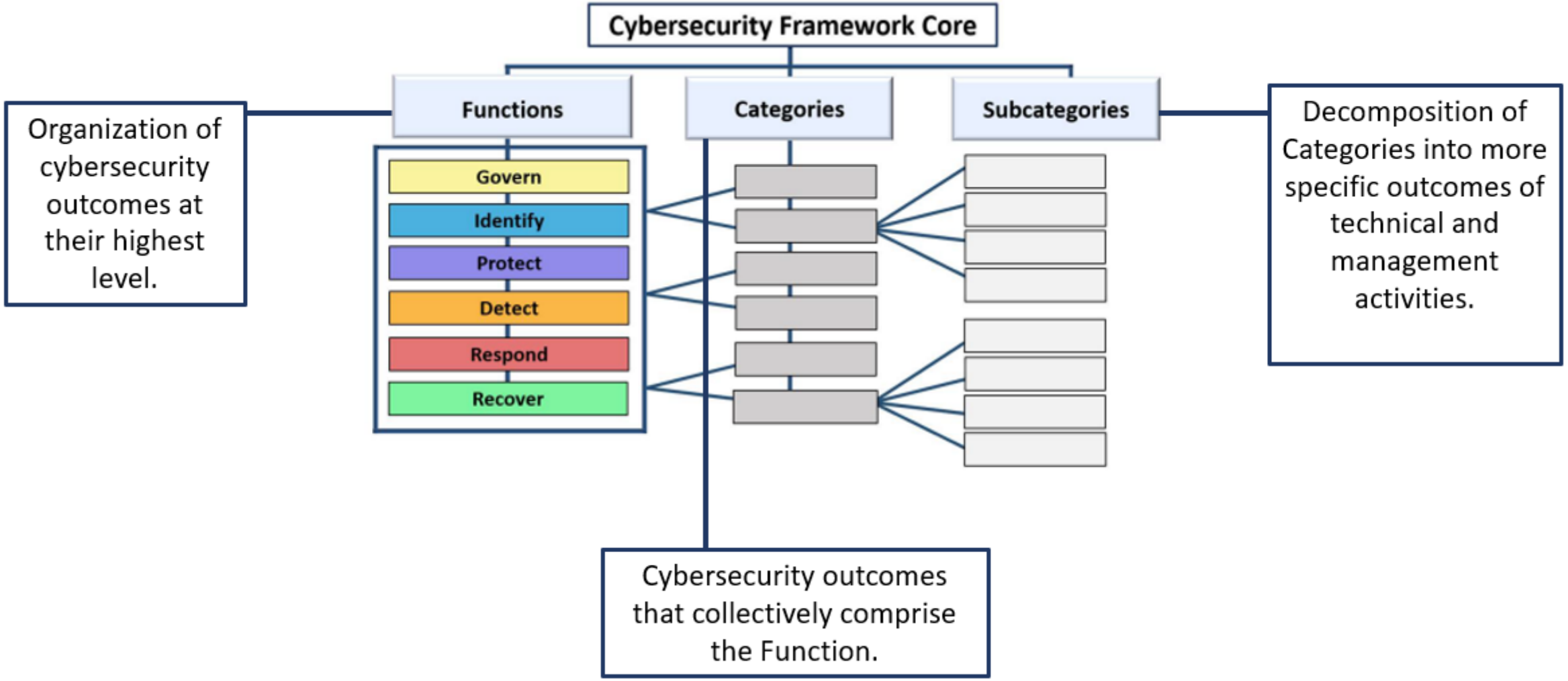
Help characterize the context and rigor of an organization's cybersecurity risk governance and management practices

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: Thwarting AI-enabled Cyber Attacks

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Examples of Impacted Activities for Thwarting AI-enabled Threats

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Example #1:

Awareness and Training (PR.AT) With advancements in the use of AI by cyber adversaries, organizations may need to consider revising their employee training (PR.AT-01) to raise awareness of AI-enabled spear phishing methods or other social engineering attacks. An organization may need to add new training for staff in specialized cybersecurity roles (PR.AT-02) as the organization adopts AI technologies.

Example #2:

Risk Assessment (ID.RA) AI technologies are used by our adversaries to become faster and more efficient at exploiting a vulnerability from the time it is discovered. Organizations may need to revisit their current processes for how they identify vulnerabilities (ID.RA-01) or for responding to vulnerability disclosures (ID.RA-08).

How AI is Changing the Adversary's Cyber Attack Game

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

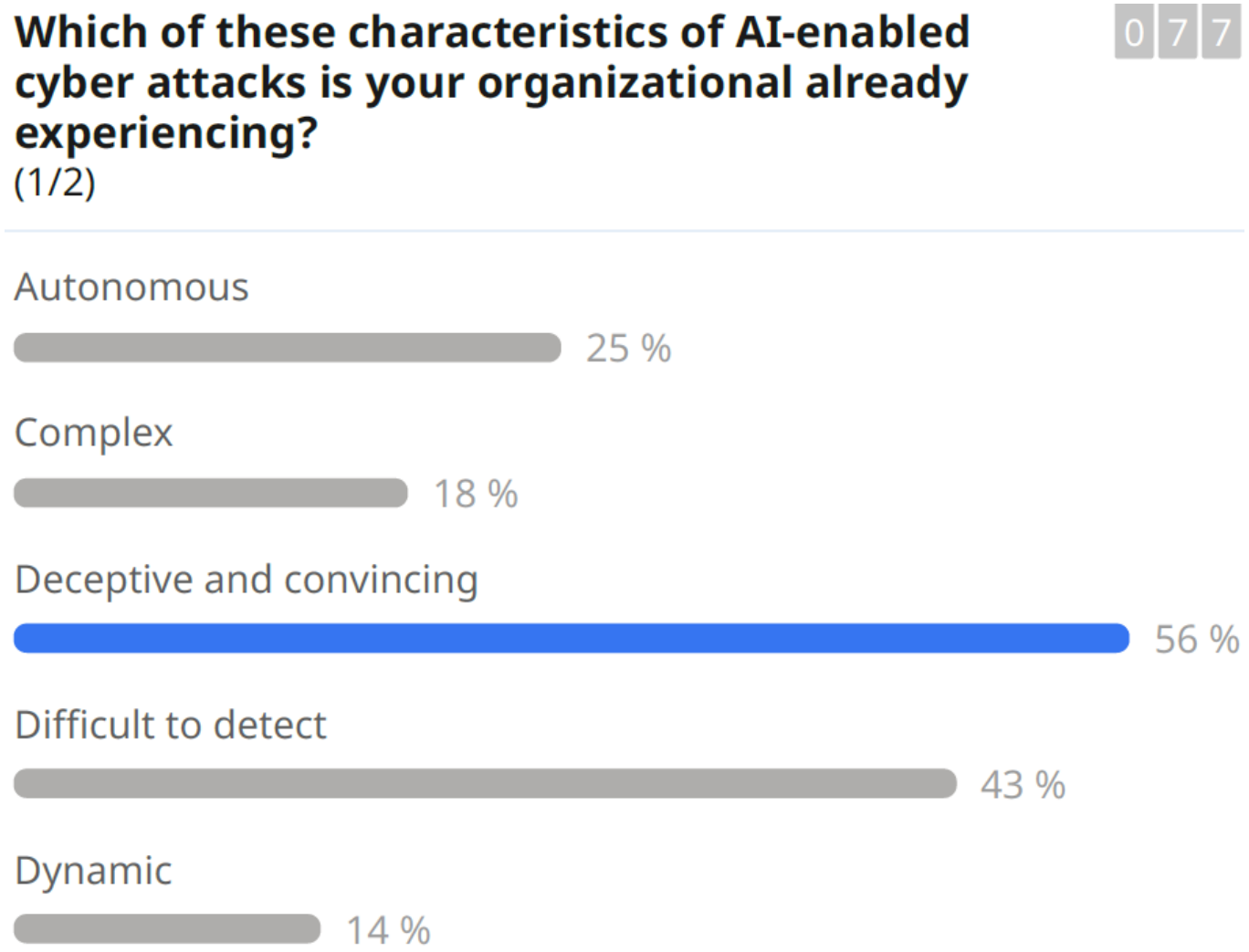
Accelerating | Optimizing | Automating | Scaling

AI-enabled cyber attacks can be:

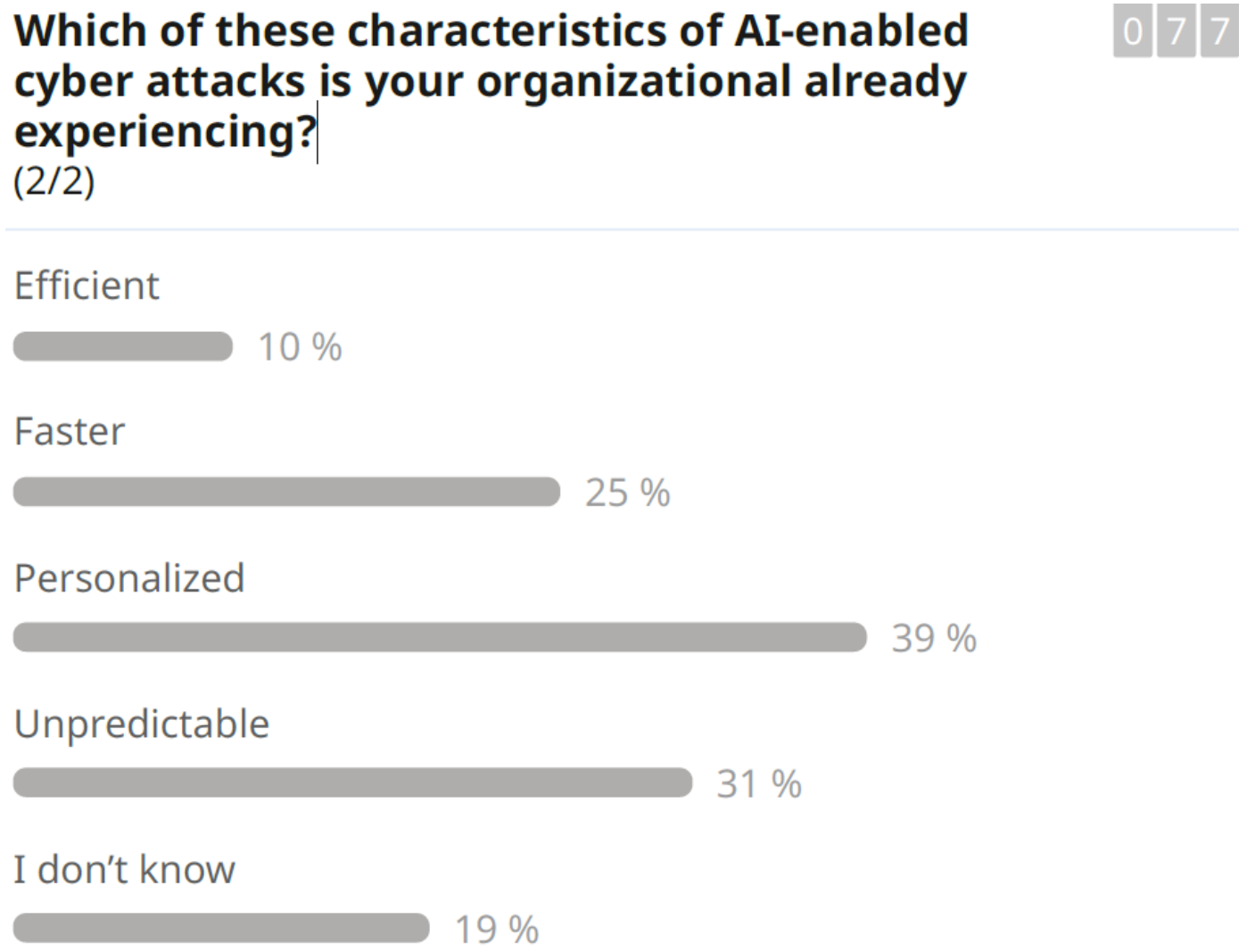
Faster	Efficient	Autonomous
Personalized	Deceptive and convincing	Complex
Dynamic	Difficult to detect	Unpredictable

Slido.com
#CyberAI_WS3

Slido Results: *How AI is Changing the Adversary's Cyber Attack Game (1 of 2)*



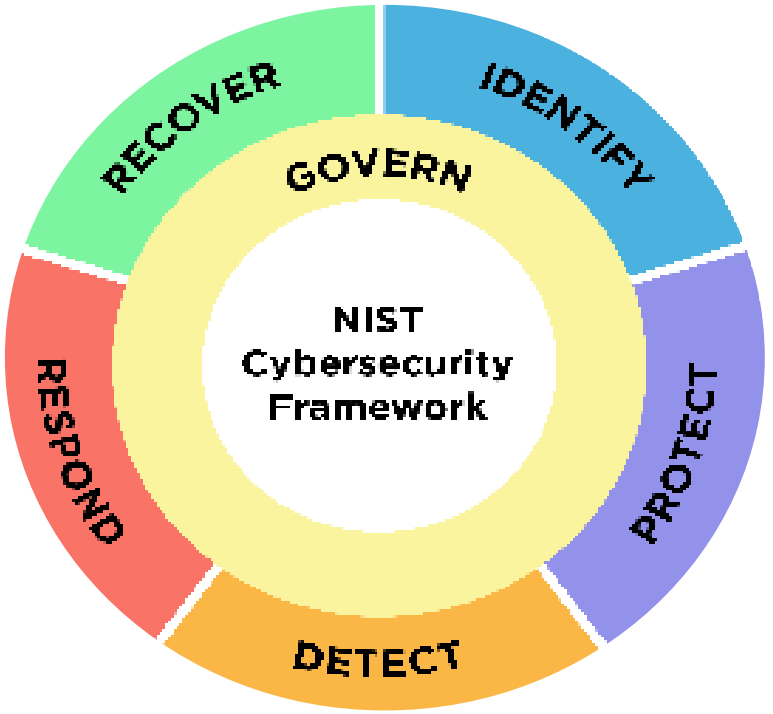
Slido Results: *How AI is Changing the Adversary's Cyber Attack Game (2 of 2)*



Impacts of Thwarting AI-enabled Cyber Attacks on CSF Functions

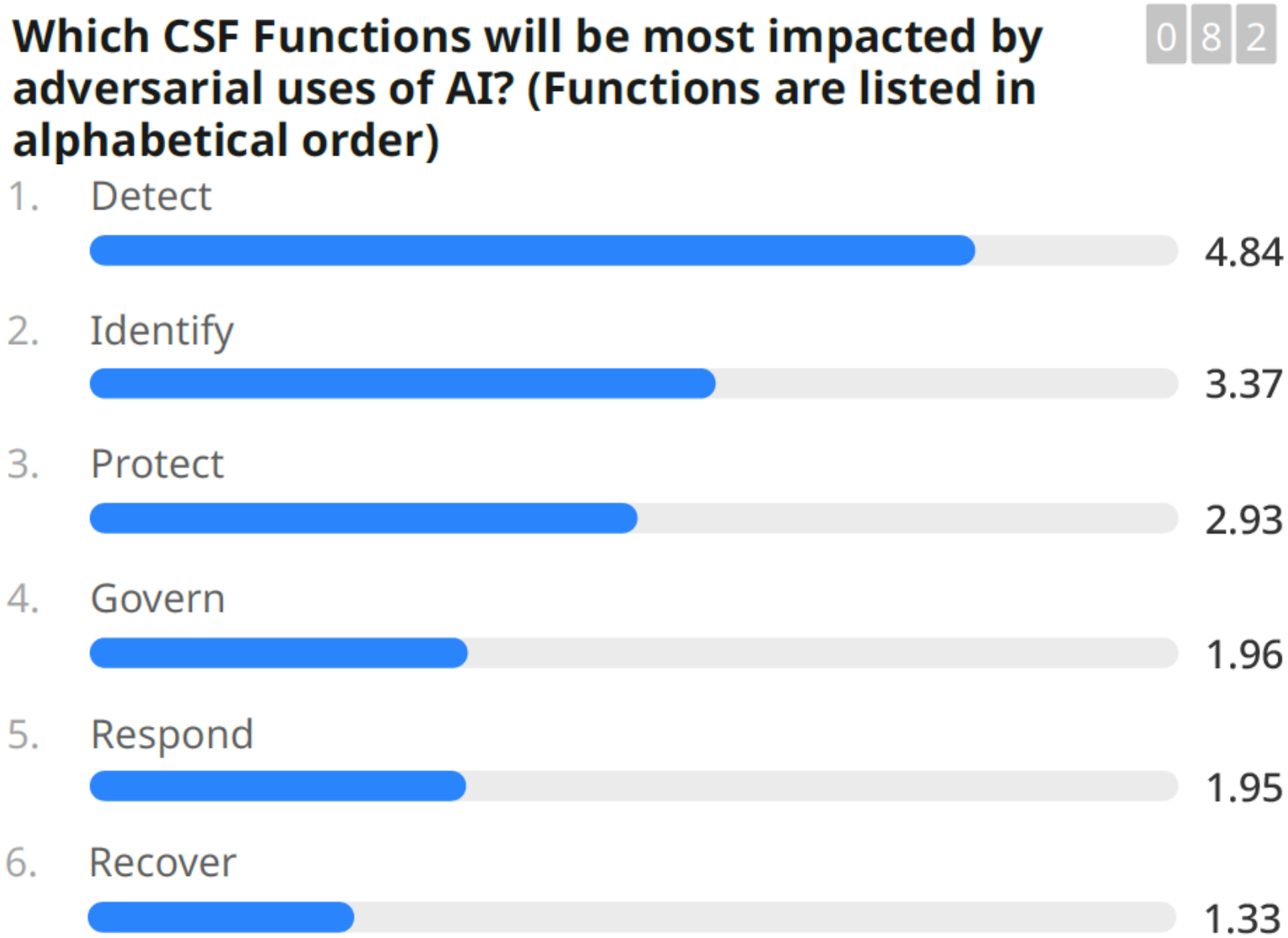
Which CSF Functions will be most impacted by adversarial uses of AI?

The poll results will inform the order in which we discuss the Categories for each Function.



Slido.com
#CyberAI_WS3

Slido Results: *Impacts of Thwarting AI-enabled Cyber Attacks on CSF Functions*



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™ • 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
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)	0.29
Risk Management Strategy (GV.RM)	0.04
Roles, Responsibilities, and Authorities (GV.RR)	0.21
Policy (GV.PO)	0.00
Oversight (GV.OV)	0.00
Cybersecurity Supply Chain Risk Management (GV.SC)	0.46

IDENTIFY	Heatmap
Asset Management (ID.AM)	0.67
Risk Assessment (ID.RA)	1.00
Improvement (ID.IM)	0.46

PROTECT	Heatmap
Identity Management, Authentication and Access Control (PR.AA)	0.83
Awareness and Training (PR.AT)	0.08
Data Security (PR.DS)	0.96
Platform Security (PR.PS)	0.79
Technology Infrastructure Resilience (PR.IR)	0.38

DETECT	Heatmap
Continuous Monitoring (DE.CM)	0.88
Adverse Event Analysis (DE.AE)	0.63

RESPOND	Heatmap
Incident Management (RS.MA)	0.17
Incident Analysis (RS.AN)	0.08
Incident Response Reporting and Communication (RS.CO)	0.17
Incident Mitigation (RS.MI)	0.00

RECOVER	Heatmap
Incident Recovery Plan Execution (RC.RP)	0.04
Incident Recovery Communications (RC.CO)	0.00

FOR DISCUSSION PURPOSES ONLY

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

Detect

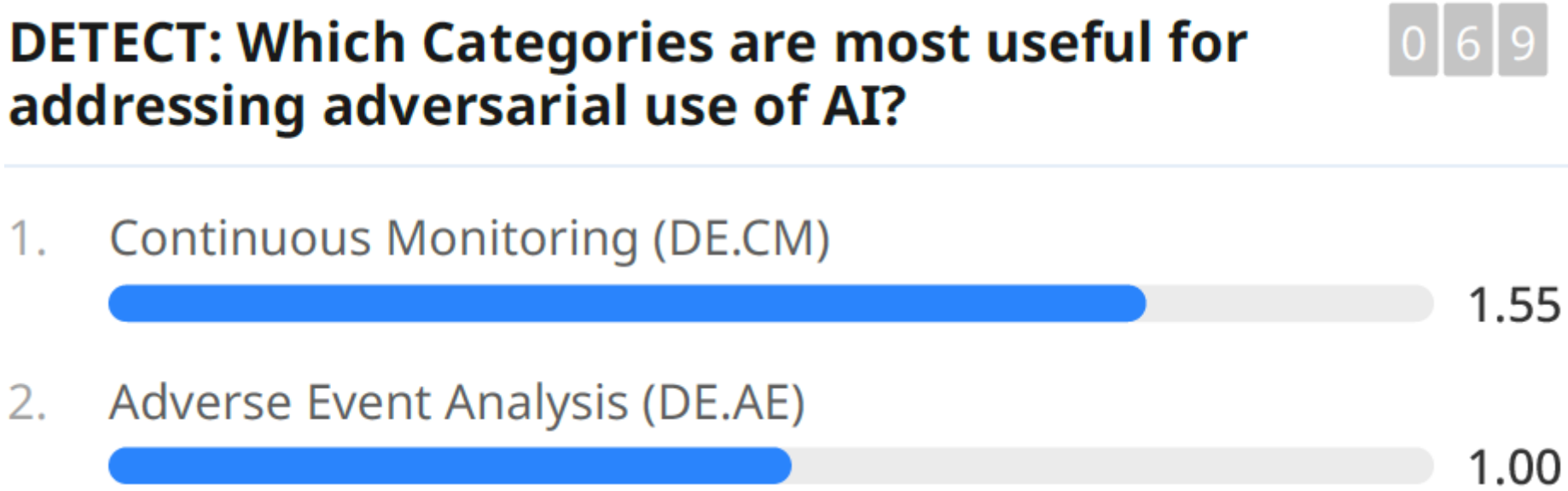
Slido.com
#CyberAI_WS3

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.	0.63
Continuous Monitoring (DE.CM)	Assets are monitored to find anomalies, indicators of compromise, and other potentially adverse events.	0.88

FOR DISCUSSION PURPOSES ONLY

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

Slido Results: *Detect*



Identify

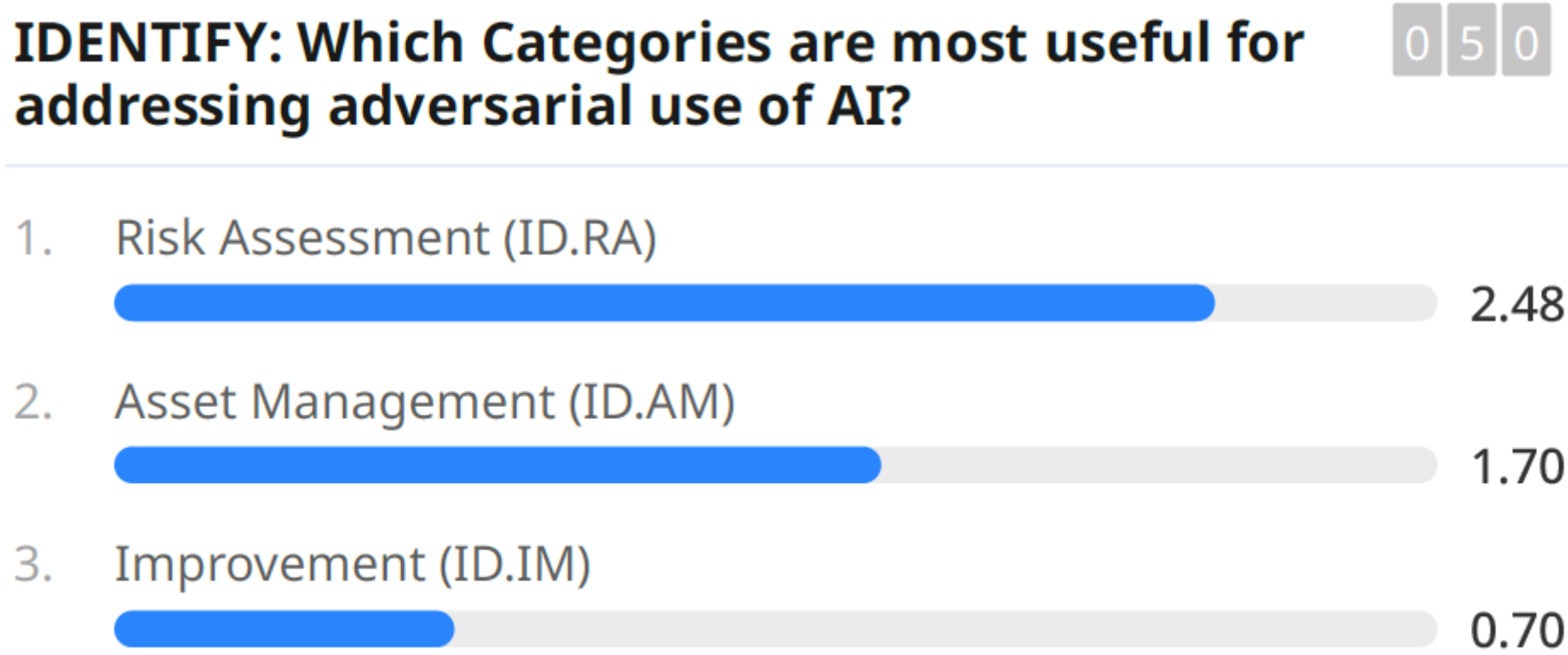
Slido.com
#CyberAI_WS3

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.	0.67
Improvement (ID.IM)	Improvements to organizational cybersecurity risk management processes, procedures and activities are identified across all CSF Functions.	0.46
Risk Assessment (ID.RA)	The cybersecurity risk to the organization, assets, and individuals is understood by the organization.	1.00

FOR DISCUSSION PURPOSES ONLY

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

Slido Results: *Identify*



Protect

Slido.com
#CyberAI_WS3

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.	0.08
Data Security (PR.DS)	Data are managed consistent with the organization's risk strategy to protect the confidentiality, integrity, and availability of information.	0.96
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.	0.83
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.	0.79
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.	0.38

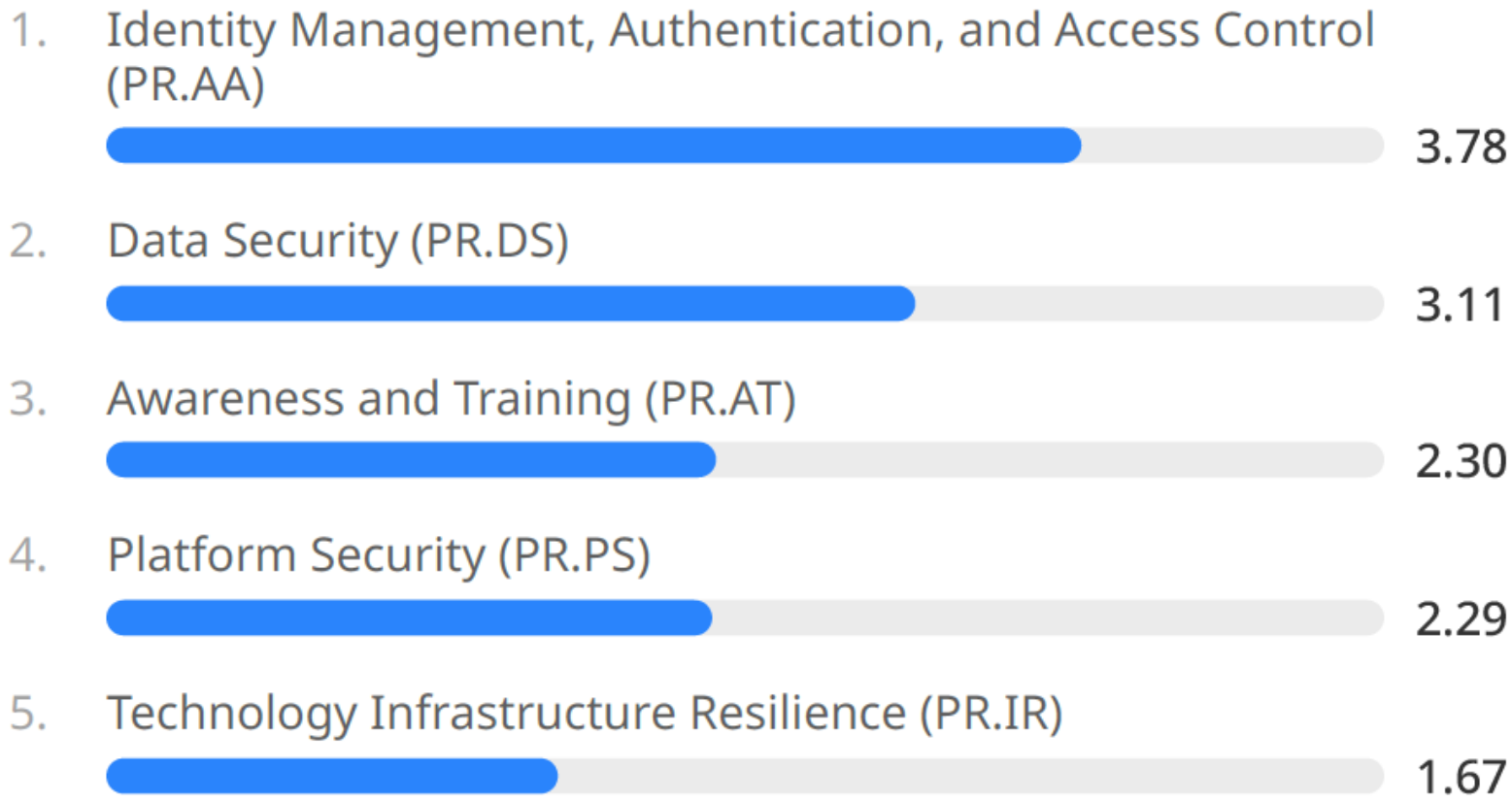
FOR DISCUSSION PURPOSES ONLY

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

Slido Results: *Protect*

PROTECT: Which Categories are most useful for addressing adversarial use of AI?

0 6 3



Respond

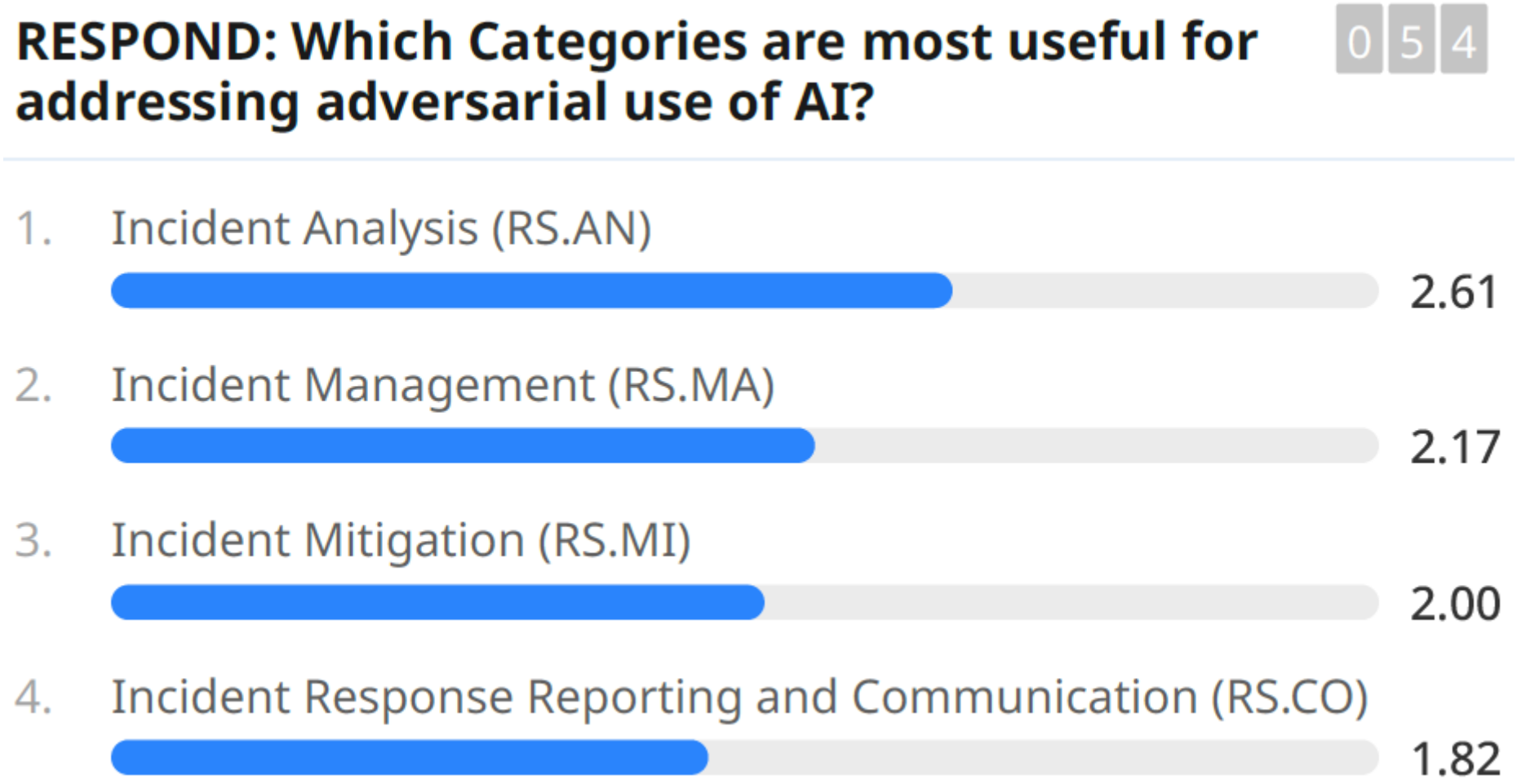
Slido.com
#CyberAI_WS3

Category	Description	Heatmap
Incident Analysis (RS.AN)	Investigations are conducted to ensure effective response and support forensics and recovery activities.	0.08
Incident Management (RS.MA)	Responses to detected cybersecurity incidents are managed.	0.17
Incident Mitigation (RS.MI)	Activities are performed to prevent expansion of an event and mitigate its effects.	0.00
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 emphasis/potential priority):

Slido Results: *Respond*



Govern

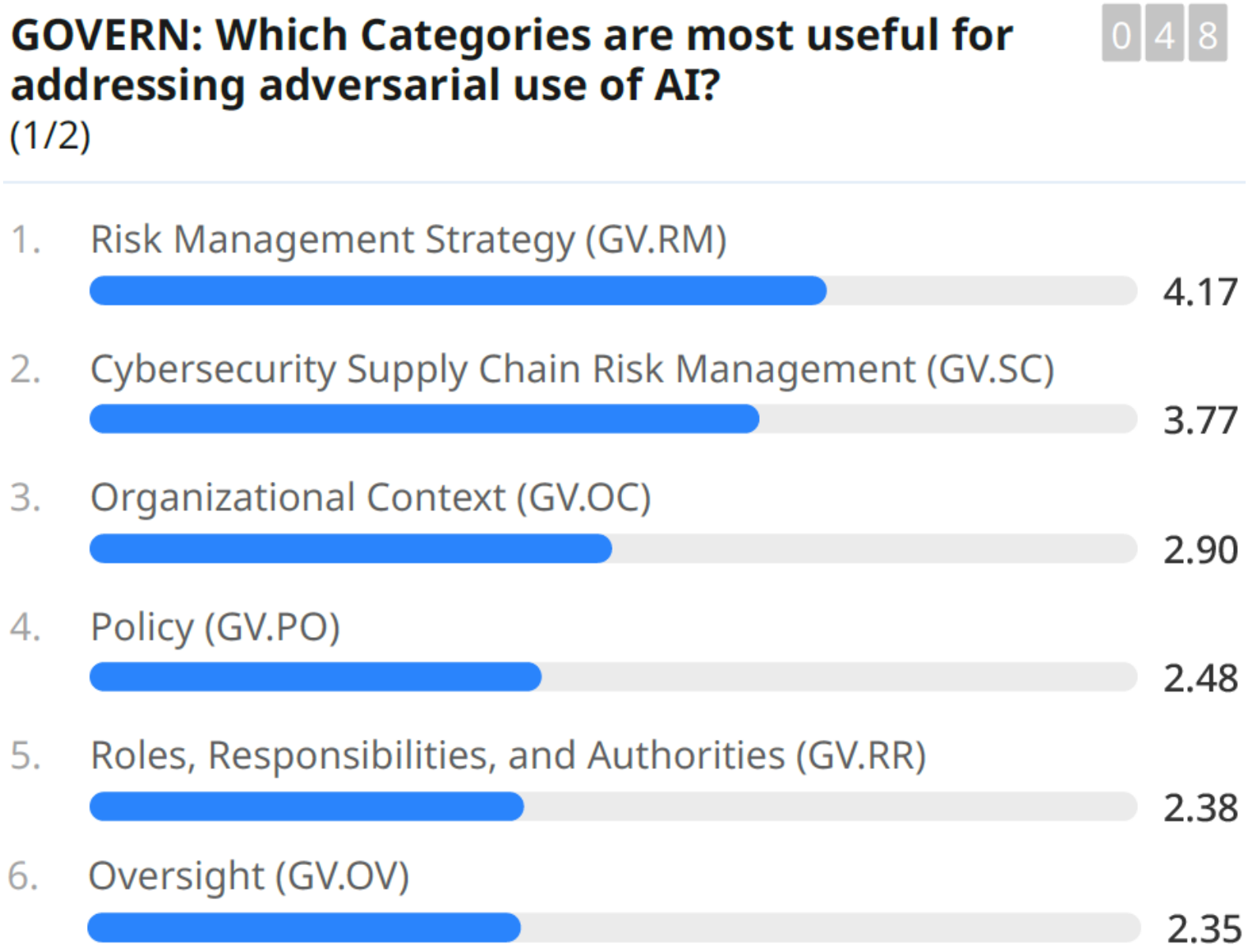
Slido.com
#CyberAI_WS3

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	0.46
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	0.29
Oversight (GV.OV)	Results of organization-wide cybersecurity risk management activities and performance are used to inform, improve, and adjust the risk management strategy	0.00
Policy (GV.PO)	Organizational cybersecurity policy is established, communicated, and enforced	0.00
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	0.04
Roles, Responsibilities, and Authorities (GV.RR)	Cybersecurity roles, responsibilities, and authorities to foster accountability, performance assessment, and continuous improvement are established and communicated	0.21

FOR DISCUSSION PURPOSES ONLY

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

Slido Results: *Govern*



Recover

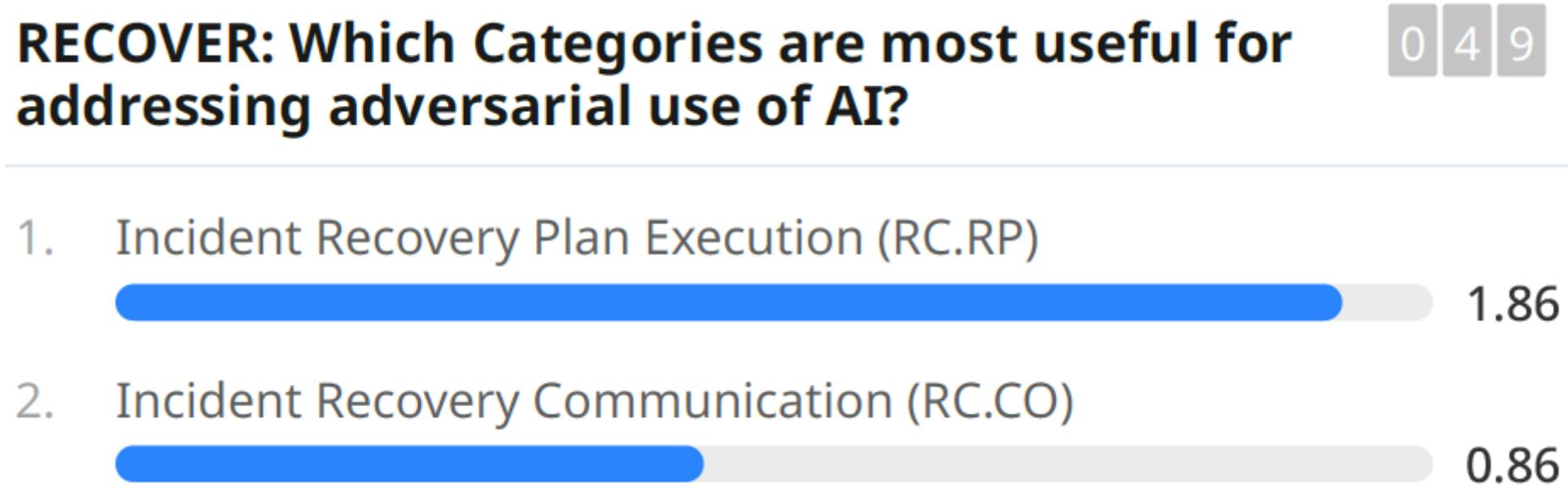
Slido.com
#CyberAI_WS3

Category	Description	Heatmap
Incident Recovery Communications (RC.CO)	Restoration activities are coordinated with internal and external parties.	0.00
Incident Recovery Plan Execution (RC.RP)	Restoration activities are performed to ensure operational availability of systems and services affected by cybersecurity incidents.	0.04

FOR DISCUSSION PURPOSES ONLY

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

Slido Results: *Recover*



Today's Focus: CSF 2.0 Categories for Thwarting AI-Enabled Cyber Attacks

- **Scope:** Bolstering defenses and building resilience to protect against new AI-enabled threat vectors.
- **Focus Area characteristics – cybersecurity considerations regarding:**
  - Understand adversarial uses of AI
  - Preparing an organization to defend against AI-enabled cyber attacks
  - Identifying and managing AI-related threats
- **Examples of AI-enabled Cyber Attacks:**
  - Self-learning/adaptive malware
  - Automated vulnerability scanning
  - Optimizing botnet coordination
  - Automating IoT activities for exploitation

Challenge Areas

- Automated and Adaptive Malware
- Phishing and Social Engineering
- Credential Stuffing and Brute Force Attacks
- AI-Driven Reconnaissance and Cyber Espionage
- Distributed Denial of Service (DDoS) Attacks
- Evasion Techniques
- Supply Chain Attacks
- AI-Powered Zero-Day Exploits
- AI-Augmented Physical Cyber Attacks
- AI-Powered Attack Automation

Slido.com
#CyberAI_WS3

Slido Results: *Describing Thwarting AI-Enabled Cyber Attacks (1 of 4)*

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

How can we better describe the scope, characteristics, and challenge areas of the Thwarting AI-enabled Cyber Attacks Focus Area? (1/4)

0 1 6

- Using both offensive and defensive AI to automate cyber-defensive strategies.
- Understanding AI and LLM related stuff in this interesting field. And to spread the knowledge
- Be able to maintain visibility & degrade/slowdown/counter-attack. Nail down your supply chain before suffering their unattended Risk. Grow your own cybersecurity AI brainiacs (aka CoE) Minimize dependency on AI Lookout for growth of SI (synthetic)
- Possible Nuance- Insider risk??
- AI Safety emphasis in these various areas to prevent attacks. How testing of AI enabled systems is implemented to help with recovery
- Understanding of the AI, ML, and LLM, like Zero Trust assume breach, limit the breach, and verify at every

Slido Results: *Describing Thwarting AI-Enabled Cyber Attacks (2 of 4)*

How can we better describe the scope, characteristics, and challenge areas of the Thwarting AI-enabled Cyber Attacks Focus Area? (2/4)

016

stop. Business must understand the potential AI enabled attacks, all about risk management and risk appetite, threshold, and tolerance.

- Expand understanding adversarial usea of AI to include refinement by type of AI model
- More information for Management understanding and buy-in to better understand the risks and real threats not just the potential purported benefits of AI and now an AI-enabled cyber attach differ and will also need resource allocation.
- A better and ongoing understanding of the adversarial efforts to optimize AI for nefarious purposes.
- Slight wording suggestion: Revise the Scope definition to include "offense" (in addition to defense language).
- Learn to be conferrable with the unconvertable of AI attacks.

Slido Results: *Describing Thwarting AI-Enabled Cyber Attacks (3 of 4)*

How can we better describe the scope, characteristics, and challenge areas of the Thwarting AI-enabled Cyber Attacks Focus Area?
(3/4)

0 1 6

- If we get into a scenario where mass-scale attacks can be conducted without time to respond and remediate, impacting multiple or many orgs at the same time.
- I see that the biggest issues associated with AI-enabled attacks is that AI can allow more targeted attacks (better phishing, zero-day vulnerability discovery, etc) as well as it can help raise the overall rate of attack (although I see near constant attacks on my systems and this rate has not really changed with AI). It really shows that lowering the attack surface (both inside and outside the firewall) is really important.
- Continuous learning and continuous scanning. Learn to know the "unknowns"
- A recommendation is to consider external threat correlation and analysis.
- There may need to

Slido Results: *Describing Thwarting AI-Enabled Cyber Attacks (4 of 4)*

How can we better describe the scope, characteristics, and challenge areas of the Thwarting AI-enabled Cyber Attacks Focus Area?
(4/4)

0 1 6

be additional subcategories to identify the nuances that are involved with AI.

- Understand the cadence required to keep up with evolution speed of AI, and AI threats
- What is done by computer systems especially if AI is compromised giving you inaccurate information as the results of an attack. If you have, had or plan for an attach put a redundant it system in for a switch over

for recovery plan execution but the communication is ongoing. 2nd system, shut the system down shut down the main system investigate then turn on the backup . Effectively terminating the hackers/ai though the ai might be a litter tougher.

Close-out

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_WS3

Slido Results: *Close-out (1 of 4)*

Close-Out Questions (1/2)

0 0 9

What additional resources (e.g., standards, mappings, tools) should we incorporate into our research?
(1/2)

- CSA AI Safety training. Updates to SaaS app testing and infrastructure. IAM may have prevented these attacks in house in the past now have more SaaS vendor exposure as access is from browser
- A list of Open-source and commercial tools, including how they map to the framework
- NIST AI RMF to NIST CSF 2.0
- Standards are essential for organizations to ensure they are utilizing tools in an efficient manner. Look to guidance from DHS CISA/NSA for additional recommendation on adversarial detection methods.
- COSAiS, deep overviews and understanding of AI, ML, and LLM. C-Suite leaders and leadership in general

Slido Results: *Close-out (2 of 4)*

Close-Out Questions (1/2)

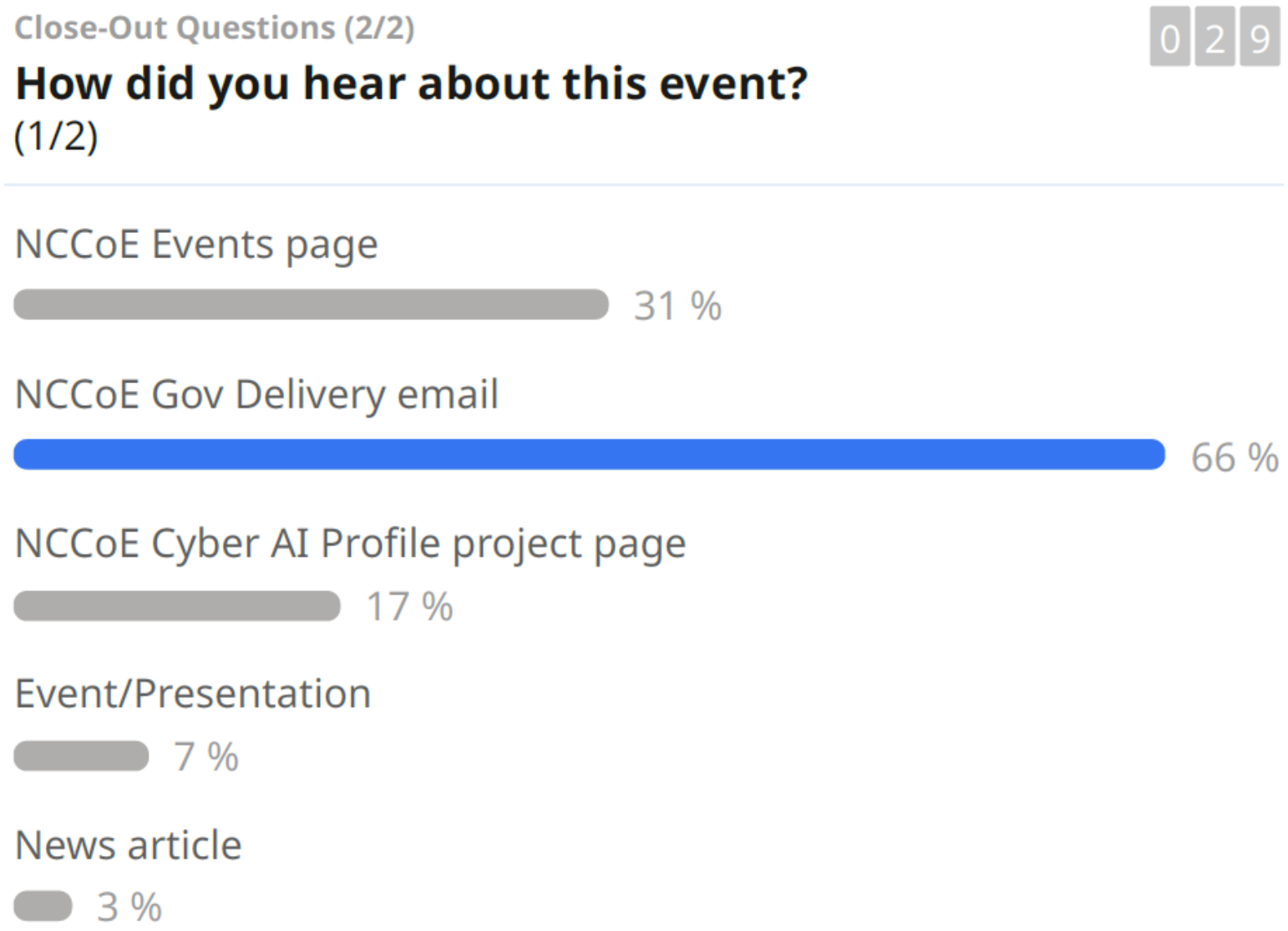
0 0 9

What additional resources (e.g., standards, mappings, tools) should we incorporate into our research?
(2/2)

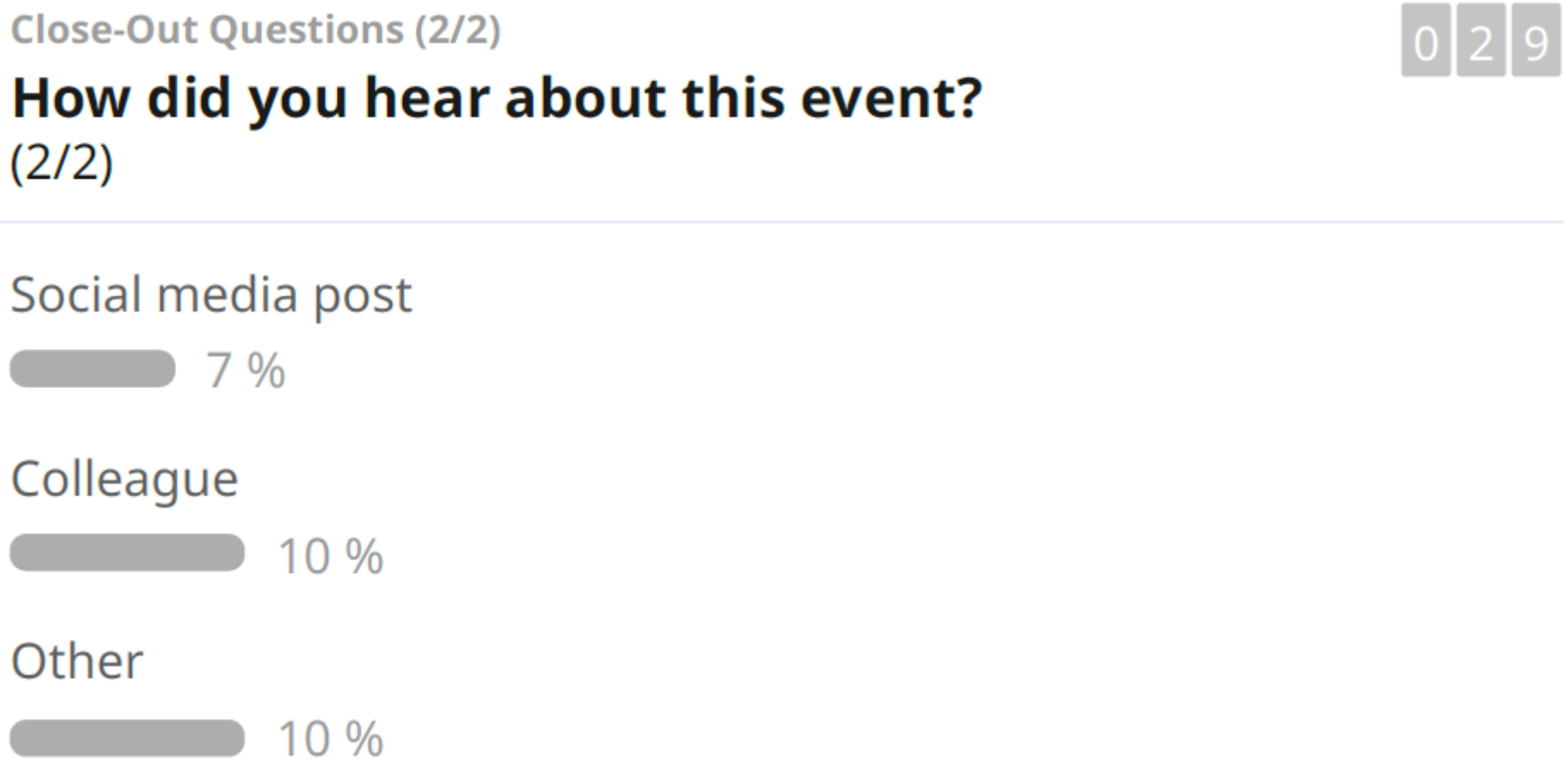
awareness and understanding or AI everything in their environments. Education and awareness will be key.

- CISA, MITRE, Pillars, Gartner, Lockheed. SAP, Google.
- Mitigation, mapping templates etc. 🙌
- consider recommendations based on some projections and forecasts of possible uplift or exponentiation of malicious cyber activity through AI enablement.
- N/A

Slido Results: *Close-out (3 of 4)*



Slido Results: *Close-out (4 of 4)*



Working Session Schedule

August 19, 2025

Conducting AI-enabled Cyber Defense

August 26, 2025

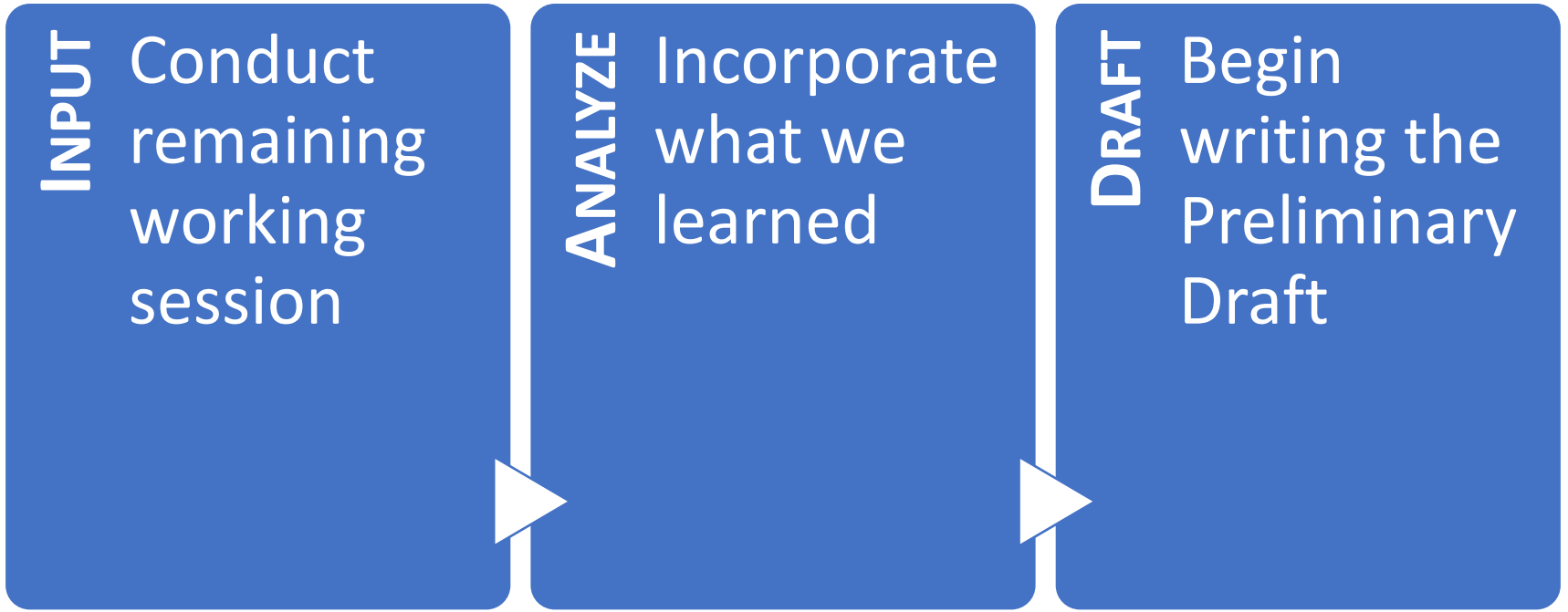
Securing AI System Components

September 2, 2025

Thwarting AI-enabled Cyber Attacks

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

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!

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

NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

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

CyberAIProfile@nist.gov

nccoe.nist.gov

@NISTcyber