Cyber AI Profile Workshop

April 3, 2025

9:00 a.m. – 5:00 p.m. EDT

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Welcome & Opening Remarks
James St. Pierre, NIST

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

NCCoE Welcome
Cherilyn Pascoe, NIST

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Welcome to the NIST NCCoE

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

A **collaborative hub** convening experts from industry, government, and academia to solve organizations' most pressing cybersecurity challenges

The NCCoE's Impact

Strengthen U.S. Cybersecurity

Provide practical guides to implement standards-based, repeatable, and scalable solutions

Improve Technology

Help vendors strengthen products' security and interoperability

Foster Public-Private Innovation

Convene industry, academia, and government to develop integrated solutions

Deliver Real-World Insights

Demonstrate solutions tested in real-world environments by leading cybersecurity experts

Support Standards Innovation

Reveal opportunities to improve standards to better address real-world challenges

Setting the Stage
Katerina Megas, NIST

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Livestream Engagement

We would love to hear from you! Please email us at CyberAIProfile@nist.gov to:

- Submit questions during Q&A
- Notify us of technical issues

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

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

- 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: An SSDF Community Profile
- 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
- Machine Learning on Privacy Enhancing Technology (PET) derived datasets
- TrojAI Challenge Rounds Based on Data Poisoning: NIST published results of the Test & Evaluation of Trojan detectors
- Automotive Cybersecurity Community of Interest (COI): Community of interest examining challenges from increased cybersecurity risk and the adoption of AI and opportunities

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:	**Outcomes:**
• 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.	• 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

Benefits of Community Profiles

Use **shared taxonomy** for cybersecurity and privacy in the context of the community

Align requirements from multiple sources

Leverage expertise across the community

Encourage **common target** outcomes

Minimize the burden by working together

Communicate about cybersecurity and privacy risk

Source (adapted): *Pascoe C, Snyder JN, Scarfone KA (2024) NIST Cybersecurity Framework 2.0: A Guide to Creating Community Profiles. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) NIST CSWP 32 ipd. https://doi.org/10.6028/NIST.CSWP.32.ipd*

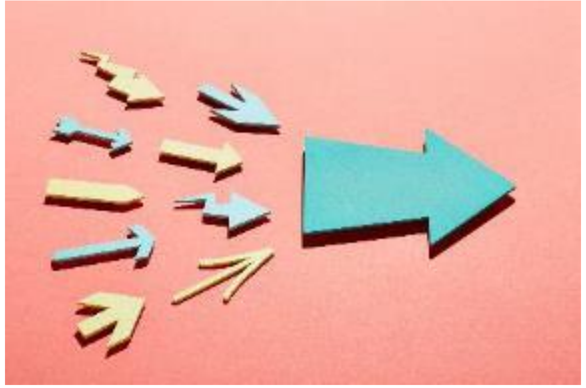
What could be in a Cyber AI Profile?

The outcomes described in the NIST Cybersecurity Framework (CSF) 2.0 provide a practical way to help organizations understand, examine, and address the cybersecurity risks introduced by the adoption of AI.

Common Priorities

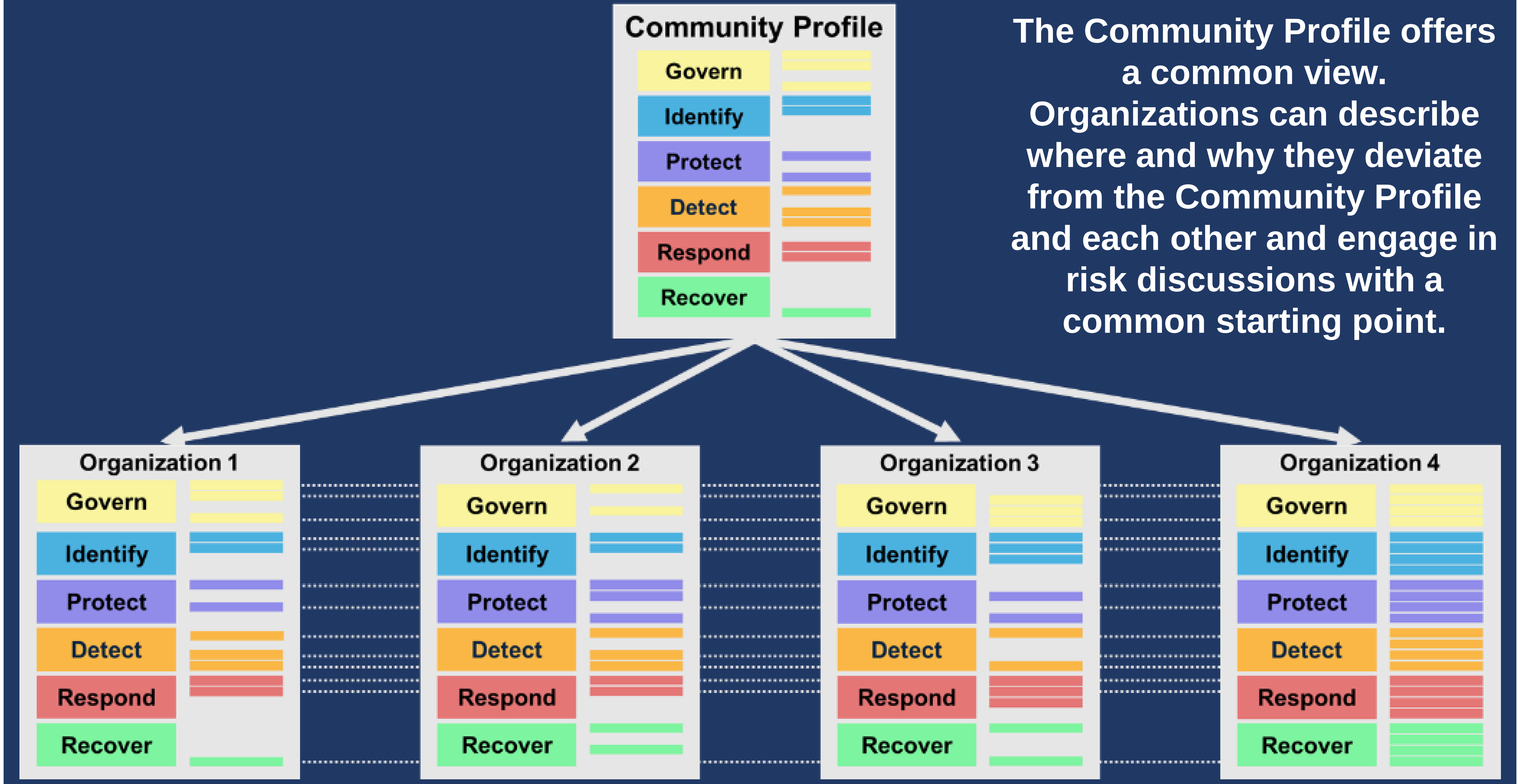
AI-specific Cybersecurity Implications

Illustrative Examples and Informative References



Mappings to Other NIST Frameworks

Community Profile Results



The Community Profile offers a common view. Organizations can describe where and why they deviate from the Community Profile and each other and engage in risk discussions with a common starting point.

Concept Paper Comment Themes

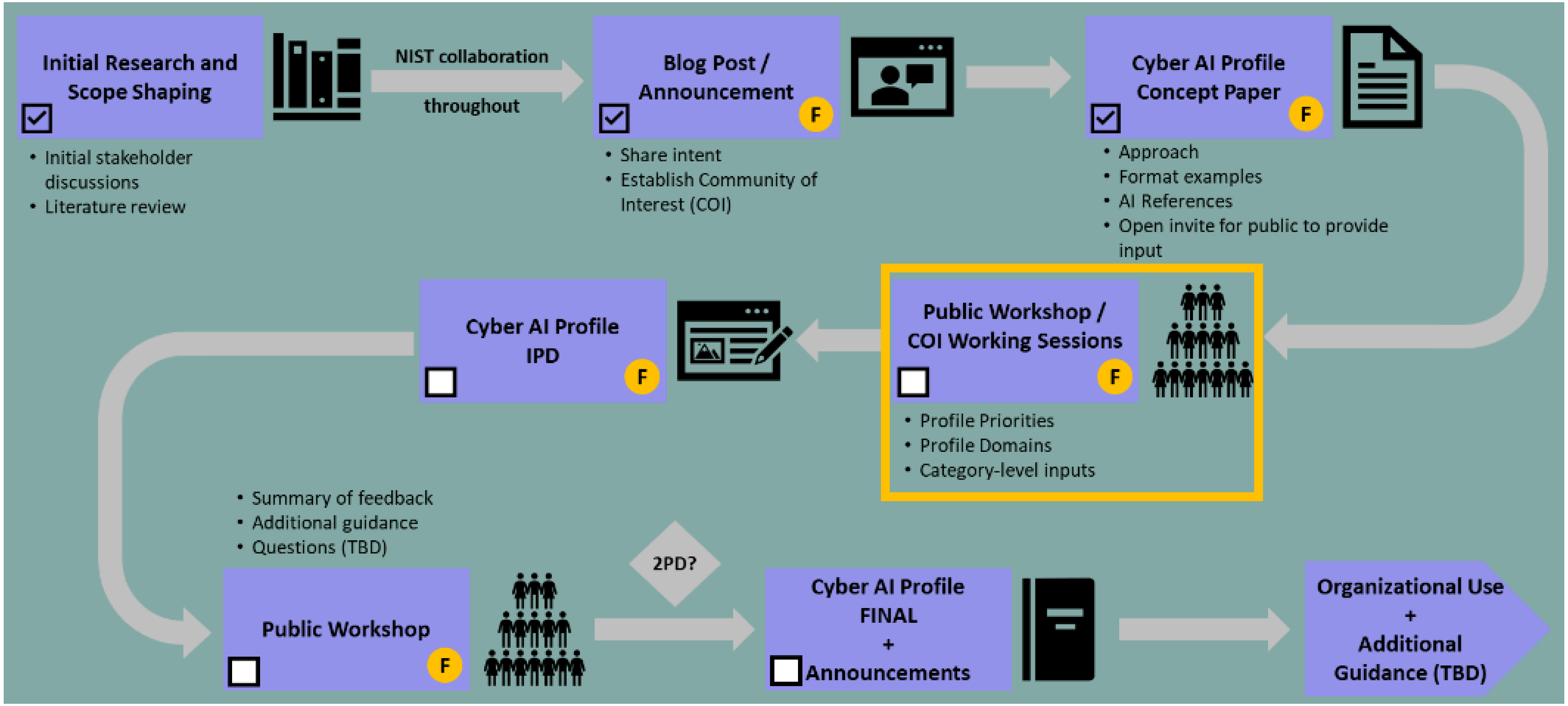
- Broad support for the approach discussed in the concept paper
  - Three focus areas
  - Starting with CSF 2.0
- AI is benefiting both the defender and the attackers
- Existing frameworks should be extended to address AI-specific risks
  - Don’t re-invent wheel
  - Interest in mapping concepts between frameworks (CSF, Privacy, and AI RMF)
  - Similar need based on RMF SP 800-53
  - Some feedback calling for integrating the AI RMF
- Data is the Crown Jewel—and the weakest link, consider data in the scope of securing AI
- AI supply chain security is the next big battleground
- Other issues on peoples’ minds:
  - Data center
  - Human-in-the-loop
  - Agentic AI
  - Governance frameworks to ensure AI solutions adhere to security best practices and compliance mandates.
- Practical guidance on implementing the Cyber AI Profile, including training, tools, and resources.
- Differentiate between AI-augmented and AI-native attacks

Workshop Goals

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

- Solidify direction of the Cyber AI Profile
- Clarify what we heard in the concept paper comments
- Hear more from YOU regarding insights, experiences, and considerations to further inform the discussions in the Cyber AI Profile

Cyber AI Profile Roadmap



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

What to Expect Today

- **Remainder of this morning (hybrid):**
  - Fireside chat to explore how we should apply the CSF and AI RMF for cybersecurity and AI
  - Panels regarding:
    - Protection of AI Systems
    - Defensive and Adversary Use of AI

- **This afternoon (in-person only):**
  - Facilitated breakout sessions to further explore questions in the concept paper and public comments
  - Three tracks:
    - A: Focus Area Descriptions
    - B: Anticipated Profile Uses and Elements
    - C: Priorities in the CSF Core

Fireside Chat: Potential Interplays Between the CSF and AI RMF

Moderator:

Dan Caprio, DLA Piper

Panelist:

Stephen Quinn, NIST

Panelist:

Martin Stanley, NIST

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Panel: Protection of AI Systems

Moderator:
Victoria Pillitteri, NIST

Panelist:
Johann Dettweiler, stackArmor

Panelist:
Faisal Khan, Protect AI

Panelist:
Arun Pamulapati, Databricks

Panelist:
Charley Snyder, Google

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Panel: Defensive and Adversary Use of AI

Moderator:
Martin Stanley, NIST

Panelist:
Drew Bagley, CrowdStrike

Panelist:
Dan Kent, Cloudflare

Panelist:
Michelle Sahar, OpenPolicy

Panelist:
Rob Sandler, Trend Micro

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Morning Wrap-up and Afternoon Breakout Session Plans

Katerina Megas, NIST

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

THANK YOU

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

CyberAIProfile@nist.gov

nccoe.nist.gov

@NISTcyber

Housekeeping

This breakout session is under **Chatham House Rule.** Participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s) may be revealed.

- Members of the press, please identify yourself and your organization
- Please raise your hand to contribute
- Please provide your name and organization prior to speaking
- Be respectful of others
- **Please silence phones**

Breakout Sessions

Please visit each track at the color-coded times below that correspond to the sticker on your badge to help us ensure there is room for everyone in each session.

Track A

Focus Area Descriptions

1:30 – 2:25 (R)
2:35 – 3:30 (Y)
3:40 – 4:35 (B)

Room 5

Track B

Anticipated Profile Uses and Elements

1:30 – 2:25 (B)
2:35 – 3:30 (R)
3:40 – 4:35 (Y)

Room 3ABC

Track C

Priorities in the CSF Core

1:30 – 2:25 (Y)
2:35 – 3:30 (B)
3:40 – 4:35 (R)

Breakout Room 3D

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Event Page & Agenda

Work Session Locations

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Room: 3D

Track C: Priorities in the CSF Core

1:30 – 2:25 (Y)
2:35 – 3:30 (B)
3:40 – 4:35 (R)

Room: 3ABC

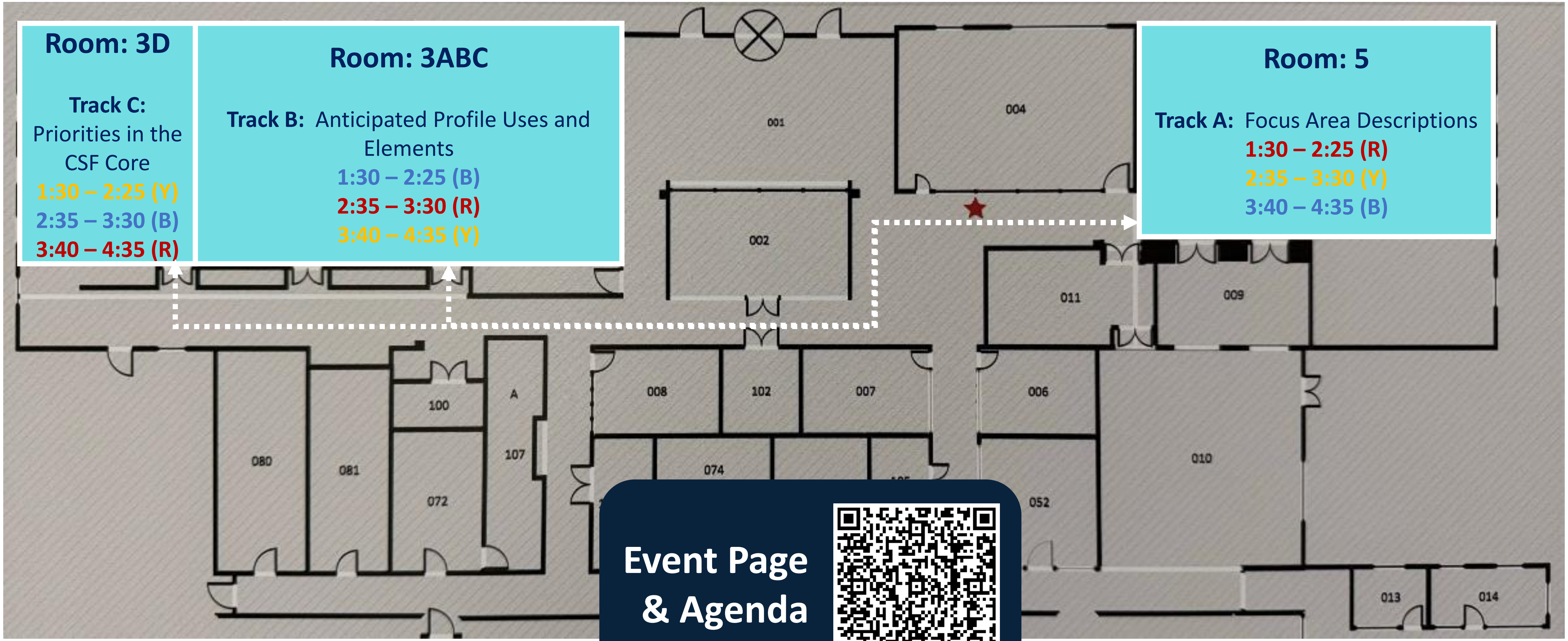
Track B: Anticipated Profile Uses and Elements

1:30 – 2:25 (B)
2:35 – 3:30 (R)
3:40 – 4:35 (Y)

Room: 5

Track A: Focus Area Descriptions

1:30 – 2:25 (R)
2:35 – 3:30 (Y)
3:40 – 4:35 (B)



Event Page & Agenda

Track A - Focus Area Descriptions

Housekeeping

This breakout session is under **Chatham House Rule.** Participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s) may be revealed.

- **Recording is prohibited**
- Members of the press, please identify yourself and your organization
- Please raise your hand to contribute
- Please provide your name and organization prior to speaking
- Be respectful of others
- **Please silence phones**

Track A - Focus Area Descriptions

The concept paper proposed three focus areas/priorities:

- Securing AI System Components
- Thwarting AI-enabled Cyber Attacks
- Using AI for Cyber-defense Activities

Questions for discussion:

- How should we describe the focus area?
- What are the key cybersecurity characteristics for each focus area?
- What resources are available today to support these areas?
- What gaps can the Profile help fill?

Securing AI System Components

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Overview: Cybersecurity risks that arise from the use of AI by organizations, including securing AI systems, components, and machine learning infrastructures, and minimizing data leakage.

- How does adopting AI expand the threat surface?
- What unique business challenges arise?
- Which system components should be covered?
- Should this discussion be separated into two parts, one for business risk and one for cybersecurity risk?

Questions for discussion:

- How should we describe the focus area?
- What are the key cybersecurity characteristics for each focus area?
- What resources are available today to support these areas?
- What gaps can the Profile help fill?

Thwarting AI-enabled Cyber Attacks

NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Overview: Determining how to defend against AI-enabled attacks.

- How is AI enabling cybersecurity adversaries?
- What can/should we do differently?
- How should we be addressing these new threat vectors?

Questions for discussion:

- How should we describe the focus area?
- What are the key cybersecurity characteristics for each focus area?
- What resources are available today to support these areas?
- What gaps can the Profile help fill?

Using AI for Cyber-defense Activities

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Overview: Assisting organizations in the use of AI with their cyber defense activities and using AI to improve privacy protections.

- How are AI-enhanced cyber capabilities changing the game?
- What risks does integrating these new capabilities introduce?
- How do we assess the efficacy of these new capabilities for cybersecurity?

Questions for discussion:

- How should we describe the focus area?
- What are the key cybersecurity characteristics for each focus area?
- What resources are available today to support these areas?
- What gaps can the Profile help fill?

Additional Focus Areas?

- Are there other important focus areas at the intersection of cybersecurity and AI that should be considered? If so:
  - What needs do they address?
  - What should we call them?
  - How should we describe them?
  - What resources are available today to support them?

Open Discussion

- How do the key characteristics compare and contrast across the focus areas?
- Additional thoughts?

Track B - Anticipated Profile Uses and Elements

Housekeeping

This breakout session is under **Chatham House Rule.** Participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s) may be revealed.

- Members of the press, please identify yourself and your organization
- Please raise your hand to contribute
- Please provide your name and organization prior to speaking
- Be respectful of others
- **Please silence phones**

Track B - Anticipated Profile Uses and Elements

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

What would you like to see included as primary elements of a Cyber AI Profile?

Our topics for today:

- Anticipated uses of the Cyber AI Profile
- Meaningful ways to identify desired outcomes
- Mappings and Informative References
- Incorporating other frameworks beyond CSF

Track C – Priorities in the CSF Core

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Housekeeping

This breakout session is under **Chatham House Rule.** Participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s) may be revealed.

- Members of the press, please identify yourself and your organization
- Please raise your hand to contribute
- Please provide your name and organization prior to speaking
- Be respectful of others
- **Please silence phones**

AI Cybersecurity Threats and Mitigations

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

- **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 heatmap presented at this workshop was 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

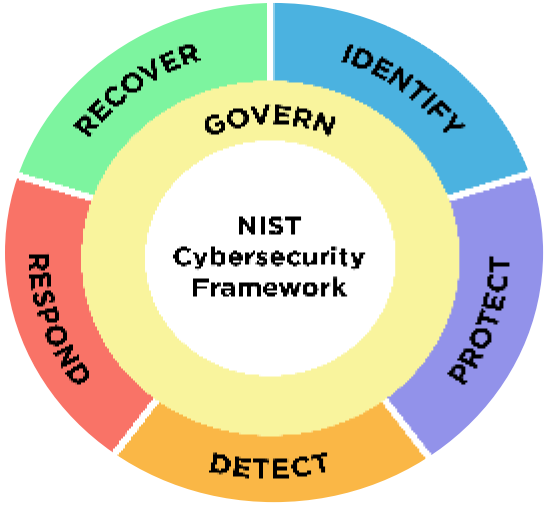
Step 1:
Examine available publications

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



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]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
ID.AM	[illegible]	[illegible]
ID.RA	[illegible]	[illegible]
ID.IM	[illegible]	[illegible]
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

FOR DISCUSSION PURPOSES ONLY

Priorities in the CSF Core

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Discussion Flow:

- Walk through each CSF Function
- Discussion questions:
  - Does the heatmap emphasize the right Categories?
  - What are the unique implications of cybersecurity and AI for the activities and outcomes in the Function?
  - What are the most critical mitigations in the Function?
  - Are there other important activities or outcomes for Cyber AI 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

Summary View

GOVERN	Heatmap
Organizational Context (GV.OC)	0.3
Risk Management Strategy (GV.RM)	0.1
Roles, Responsibilities, and Authorities (GV.RR)	0.1
Policy (GV.PO)	0.1
Oversight (GV.OV)	0.1
Cybersecurity Supply Chain [illegible] Management (GV.SC)	[illegible]

IDENTIFY	Heatmap
Asset Management (ID.AM)	0.7
Risk Assessment (ID.RA)	0.4
Improvement (ID.IM)	0.2

PROTECT	Heatmap
Identity Management, Authentication and Access Control (PR.AA)	0.5
Awareness and Training (PR.AT)	0.3
Data Security (PR.DS)	1.0
Platform Securi[illegible]	[illegible]
[illegible]structure Resilience (PR.IR)	0.5

DETECT	Heatmap
Continuous Monitoring (DE.CM)	0.4
Adverse Event Analysis (DE.AE)	0.5

RESPOND	Heatmap
Incident Management (RS.MA)	0.1
Incident Analysis [illegible]	[illegible]
[illegible] Reporting and Communication (RS.CO)	0.1
Incident Mitigation (RS.MI)	0.0

RECOVER	Heatmap
Incident Recovery Plan Execution (RC.RP)	0.0
[illegible] Communications (RC.CO)	0.0

Heatmap Legend 0-1 (degree of emphasis/potential priority): Low — Moderate — High

Govern

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	0.3
Risk Management Strategy (GV.RM)	The organization's priorities, constraints, risk tolerance and appetite statements, and assumptions are established, communicated, and used to support operati[illegible] decisions	[illegible]1
Roles, Responsibilities, and Authorities (GV.RR)	Cybersecurity roles, responsibilities, and a[illegible]ountability, performance assessment, an[illegible]ment are established and communicat[illegible]	0.1
Policy (GV.PO)	[illegible]ersecurity policy is established, communicated, and enforced	0.1
Oversigh[illegible]	Results of organization-wide cybersecurity risk management activities and performance are used to inform, improve, and adjust the risk management strategy	0.1
Cybersecurity Supply Chain Risk Management (GV.SC)	Cyber supply chain risk management processes are identified, established, managed, monitored, and improved by organizational stakeholders	0.5

Heatmap Legend 0-1 (degree of emphasis/potential priority): **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 manage[illegible] consistent with their relative importance to organizatio[illegible] organization's risk strateg[illegible]	[illegible]
Risk Assessment (ID.RA)	[illegible]ization, assets, and individuals is understood by [illegible]	0.4
Impr[illegible]ment (ID.IM)	Improvements to organizational cybersecurity risk management processes, procedures and activities are identified across all CSF Functions.	0.2

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

Low **Moderate** **High**

Protect

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 cybersecu[illegible]ing so that they can perform their cybersecurity [illegible]	0.3
Data Security (PR.DS)	Data are managed consist[illegible]on's risk strategy to protect the confidentiali[illegible]bility of information.	1.0
Platform Security (PR.PS)	[illegible] 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.	0.5

Heatmap Legend 0-1 (degree of emphasis/potential priority): Low — Moderate — High

Detect

Category	Description	Heatmap
Continuous Monitoring (DE.CM)	Assets are monitored to find anomalies, indicators of compromise, and other potentially adverse events[illegible]	[illegible]
Adve[illegible] (DE.A[illegible]	[illegible] potentially adverse events are [illegible]alyzed to characterize the events and detect cybersecurity incidents.	0.5

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

Low **Moderate** **High**

Respond

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Category	Description	Heatmap
Incident Management (RS.MA)	Responses to detected cybersecurity incidents are managed.	[illegible]
Incident Analysis (RS.AN)	Investigations are conducted to ensu[illegible] forensics and recovery activiti[illegible]	0.0
Incident Respons[illegible] and [illegible]	[illegible] are coordinated with internal and external stakeholders as required by laws, regulations, or policies.	0.1
Incident Mitigation (RS.MI)	Activities are performed to prevent expansion of an event and mitigate its effects.	0.0

Heatmap Legend 0-1 (degree of emphasis/potential priority): Low — Moderate — High

Recover

Category	Description	Heatmap
Incident Recovery Plan Execution (RC.RP)	Restoration activities are performed to ensure operational availability of [illegible] and services affected [illegible]	[illegible]
Incide[illegible] Comm[illegible]ations (RC.CO)	[illegible] with internal and external parties.	0.0

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

Low **Moderate** **High**

Workshop Closeout

Katerina Megas, NIST

Hillary Tran, MITRE

Jon Davis, MITRE

John Dombrowski, MITRE

Breakout Session Summaries

Track A

Focus Area Descriptions

Hillary Tran, MITRE

Track B

Anticipated Profile Uses and Elements

Jon Davis, MITRE

Track C

Priorities in the CSF Core

John Dombrowski, MITRE

Next Steps

NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

- Analyze what we heard during this workshop
- Identify any additional inputs needed to develop the initial public draft of the Cyber AI Profile

THANK YOU

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

CyberAIProfile@nist.gov

nccoe.nist.gov

@NISTcyber