ADVANCED CYBER SECURITY CENTER

Trusted Networks. Advancing Cyber Strategies.

ANNUAL CONFERENCE

Selected **AI** slides

Cyber Risk Governance

Leveling up in an Age of New Regulations and AI

Thank you

Strategic Partner & Lead Sponsor

ISTARI

Conference Sponsors

redsense
CYBER THREAT INTELLIGENCE

FOLEY HOAG

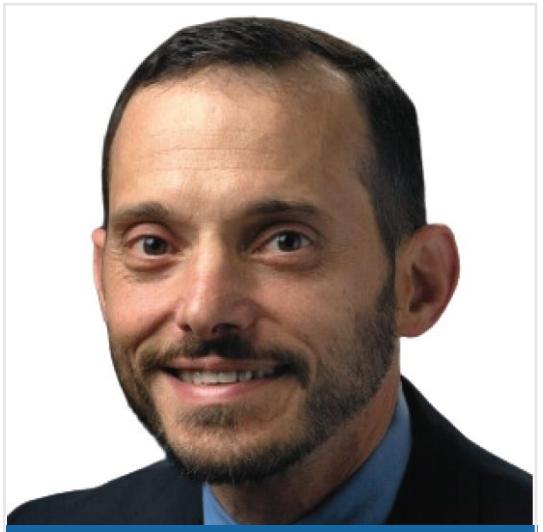
Our Host

FEDERAL RESERVE BANK OF BOSTON™

ADVANCED CYBER SECURITY CENTER

Trusted Networks. Advancing Cyber Strategies.

The AI Juggernaut and Cyber Risk Governance



Dr. Marc Zissman
MIT Lincoln Lab

The AI Juggernaut and Cyber Risk Governance

Overview

In addition to architecting and implementing AI-enabled capabilities to obtain best value, **corporate leaders must manage and mitigate risk arising from**

Adversarial threats to AI data, models, software & processes	**Compliance & ethical challenges** in an evolving AI landscape	**Threats to the organization's reputation**

It's also crucial to assess whether the AI-enabled capability is **performing as intended**

In this session, we will hear **perspectives on each of these risk management challenges** from a cross-sector panel of leaders with experience in industry and government

ACSC

The AI Juggernaut and Cyber Risk Governance

Key Session Outcomes

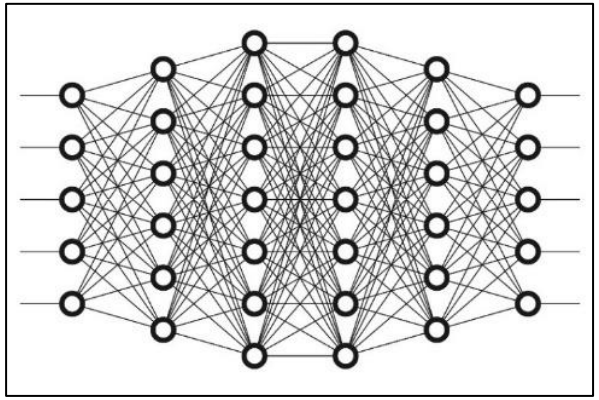
- **Case studies** from leading organizations taking advantage of AI
- Sharpened understanding of **compliance, ethics and reputational challenges** in an evolving legal landscape
- Insights into **emerging defenses** against adversarial threats
- **Toolkit** with key questions to evaluate vendors, partners, and internal systems

What has changed recently?

1. More Data

2. More Compute Power

3. Better Algorithms

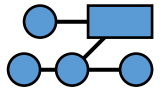


An Example AI Application: Image Recognition

Training Data
(e.g., images)

Model
Training

Model(s) of images



Training Phase

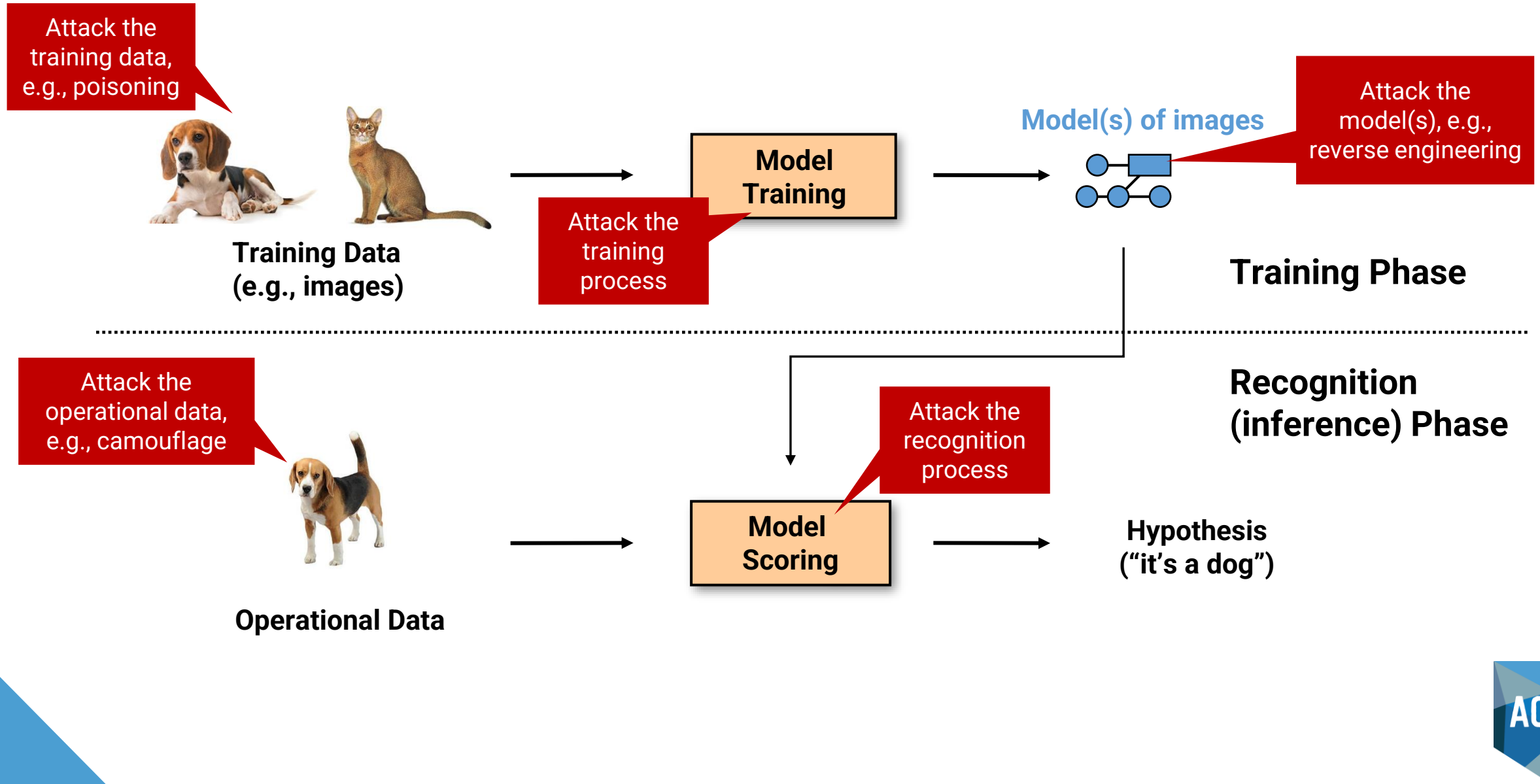
Recognition
(inference) Phase

Operational Data

Model
Scoring

Hypothesis
("it's a dog")

Attacking the Image Recognition System



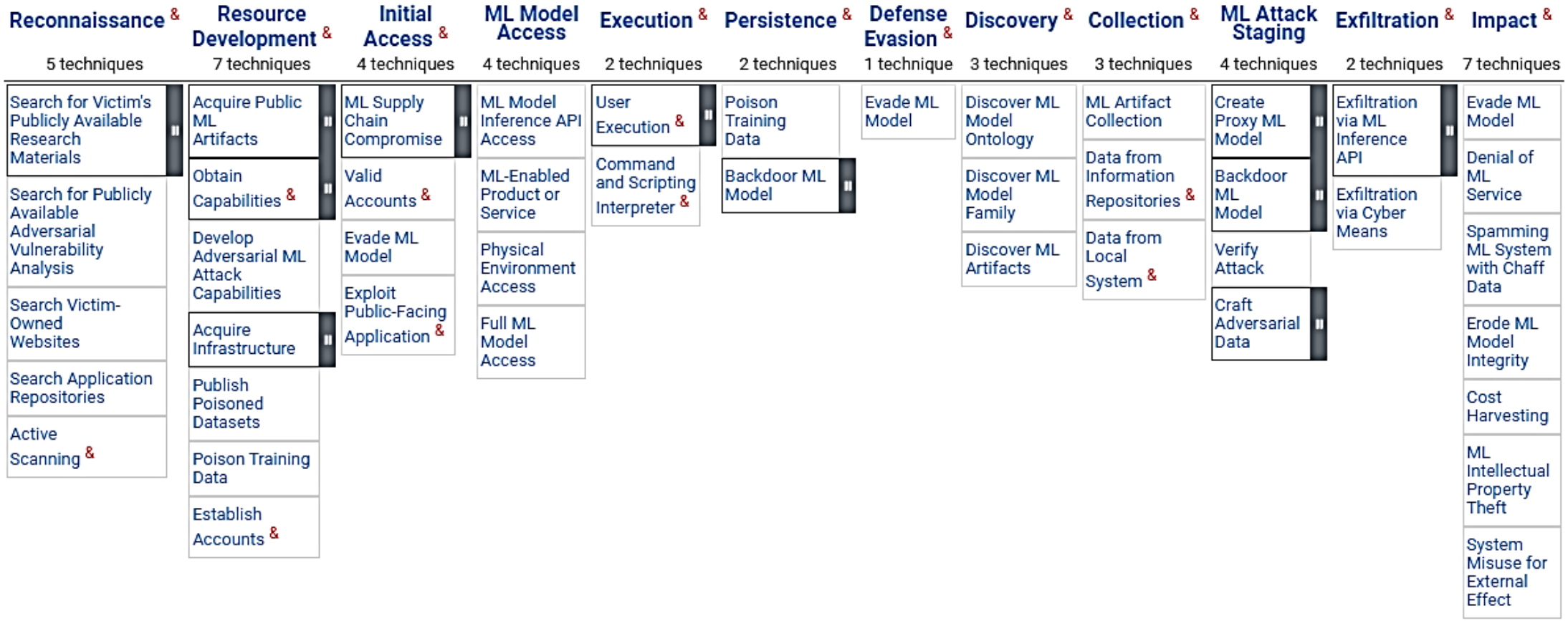
MITRE ATLAS™ Knowledge Base

MITRE ATLAS™ (Adversarial Threat Landscape for Artificial-Intelligence Systems), is a **knowledge base of adversary tactics, techniques, and case studies for machine learning (ML) systems** based on real-world observations, demonstrations from ML red teams and security groups, and the state of the possible from academic research.

ATLAS is modeled after the MITRE ATT&CK® framework and its tactics and techniques are complementary to those in ATT&CK.

For more information, see **atlas.mitre.org**

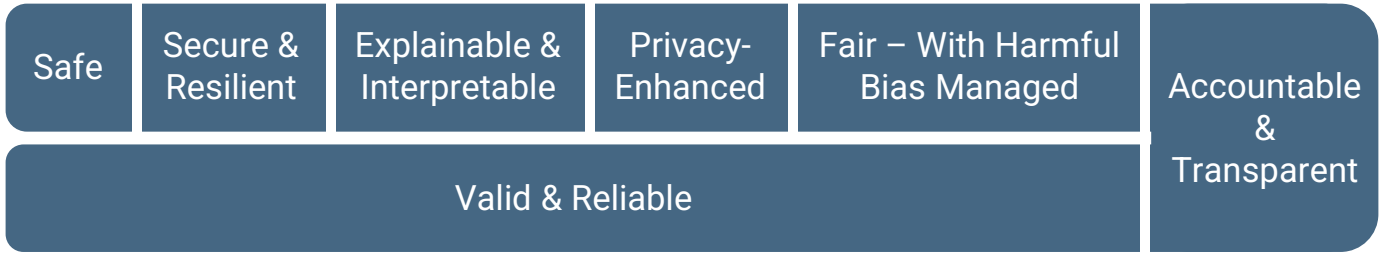
MITRE ATLAS™ Knowledge Base



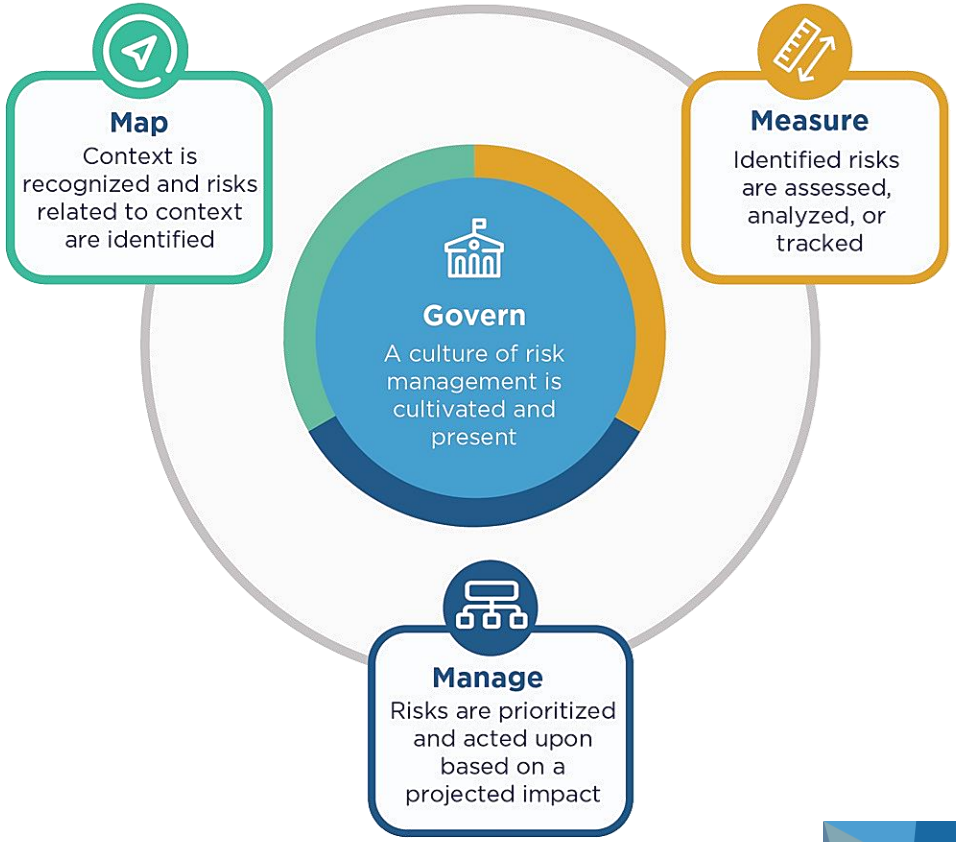
For more information, see **atlas.mitre.org**

NIST's AI Risk Management Framework (January 2023)

Characteristics of trustworthy AI systems



AI risk management functions



For more information, see **https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf**

White House Executive Order (October 2023)

- **Guiding principles for development and use of AI**
  - Safety and security, including testing & evaluations
  - Responsible innovation, competition, collaboration
  - Commitment to American workers
  - Equity and civil rights
  - Consumer protection
  - Privacy and civil liberties
  - Manage risk from government use of AI
  - Federal government should lead progress
- **Cross-agency responses: new plans, regulations, etc.**

For more information, see **https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/**

The AI Juggernaut and Cyber Risk Governance

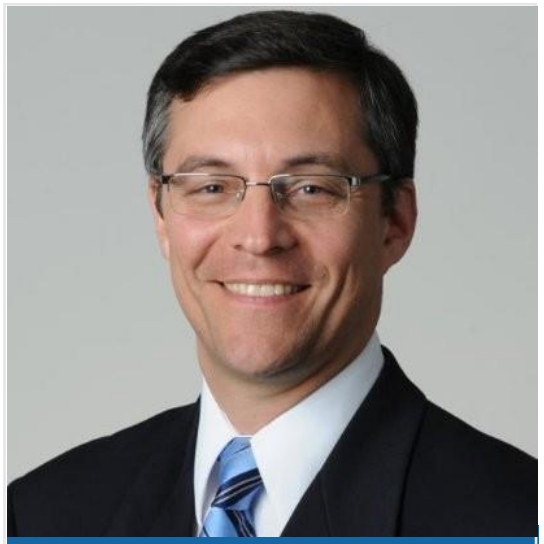
Key Session Outcomes

- **Case studies** from leading organizations taking advantage of AI
- Sharpened understanding of **compliance, ethics and reputational challenges** in an evolving legal landscape
- Insights into **emerging defenses** against adversarial threats
- **Toolkit** with key questions to evaluate vendors, partners, and internal systems

ADVANCED CYBER SECURITY CENTER

Trusted Networks. Advancing Cyber Strategies.

The AI Juggernaut and Cyber Risk Governance



Dr. Mark Maybury
Lockheed Martin

Adeel Saeed
Kyndryl

Avi Gesser
Debevoise & Plimpton

Policy and Practice Considerations for Responsible AI

- Define and establish **governance**
- Consider broad set of **relevant use cases**
- Establish and use **policies and procedures** across full life-cycle
- **Educate stakeholders** on benefits and risks
- Establish **guidelines** for appropriate and inappropriate use

Special considerations for **generative AI**

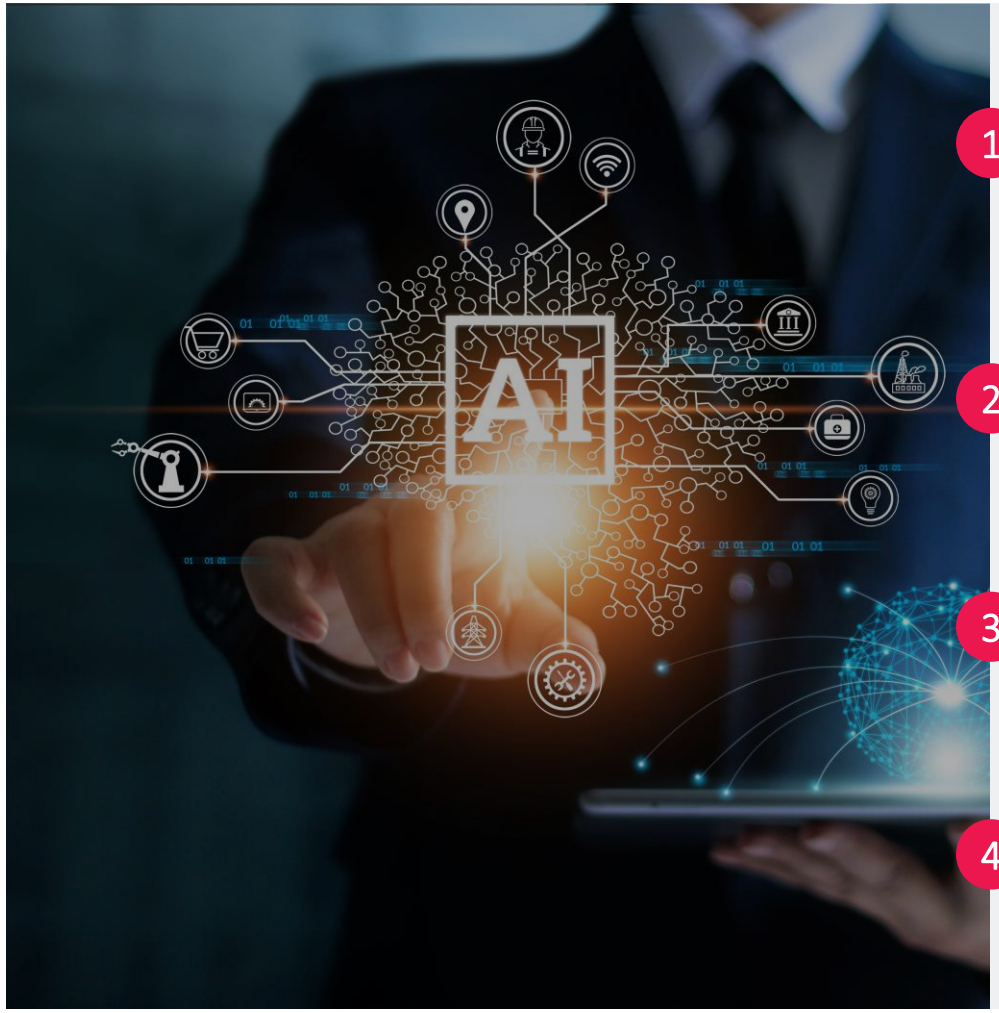
- Train users on capabilities & limitations, prompt engineering
- Train users on privacy and IP awareness
- Internally sequester large language models w/proprietary data
- Establish controls for access, model release, use monitoring, oversight

Foster AI literacy at all levels

As presented by Mark Maybury

ACSC

Examples of Cyber Risks Associated with AI



1. Moving large volumes of sensitive data from a secure on-premises location to a less secure data lake or cloud environment without proper protections
2. Sharing confidential data for training or operating the AI with a less secure AI provider or consultant.
3. Connecting GenAI Agents to emails, calendars, browsers, without sufficient security controls.
4. Policies and training to prevent phishing and BECs do not properly account for risks of Deepfake attacks.

As presented by Ave Gesser