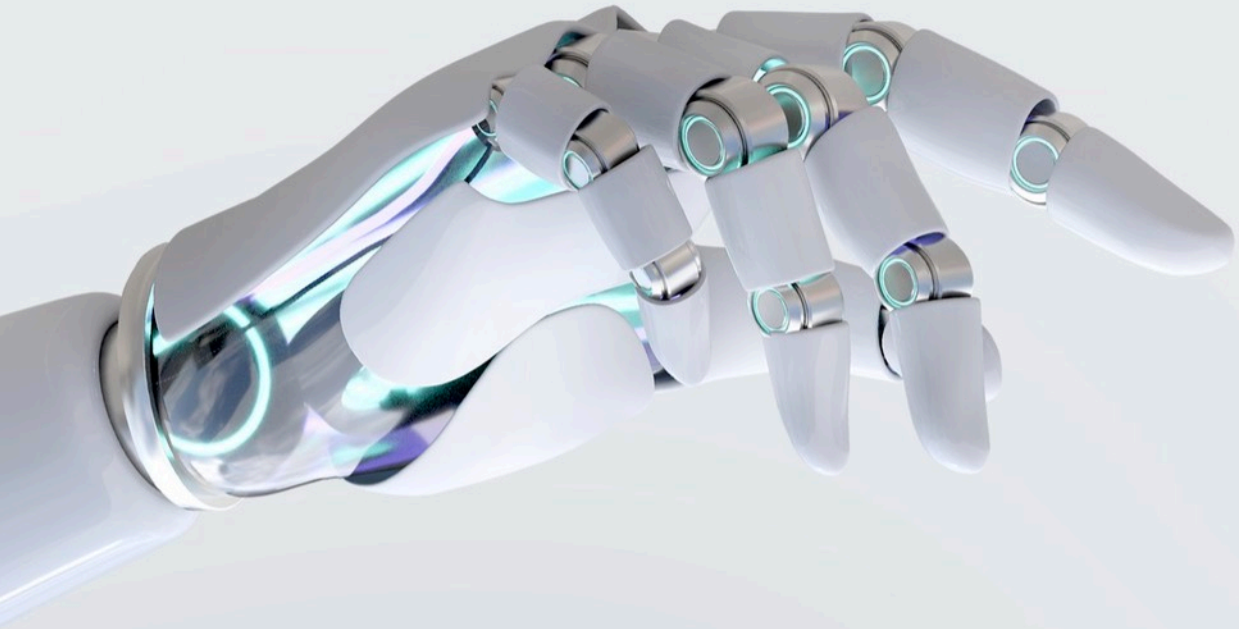


SAFEGUARDING THE FUTURE

Navigating Cybersecurity & Compliance In The Age Of Generative AI

10th Annual

New York Metro Joint Cyber Security Conference & Workshop

October 19th – 20th, 2023

InfoSecurity.NYC

The Future Of AI Is Not Just About Innovation, But Also, About Securing That Innovation

The Rapid Evolution Of The Cyber Threat Landscape And The Emergence Of Powerful Generative AI Are Compelling Cybersecurity Leaders To Continuously Adapt Protection Strategies And Collaborate Closely With Stakeholders' Enterprise-wide To Manage Emerging Risks Proactively

100%

Nearly all business leaders say their company is prioritizing at least one initiative related to AI systems in the near term

35%

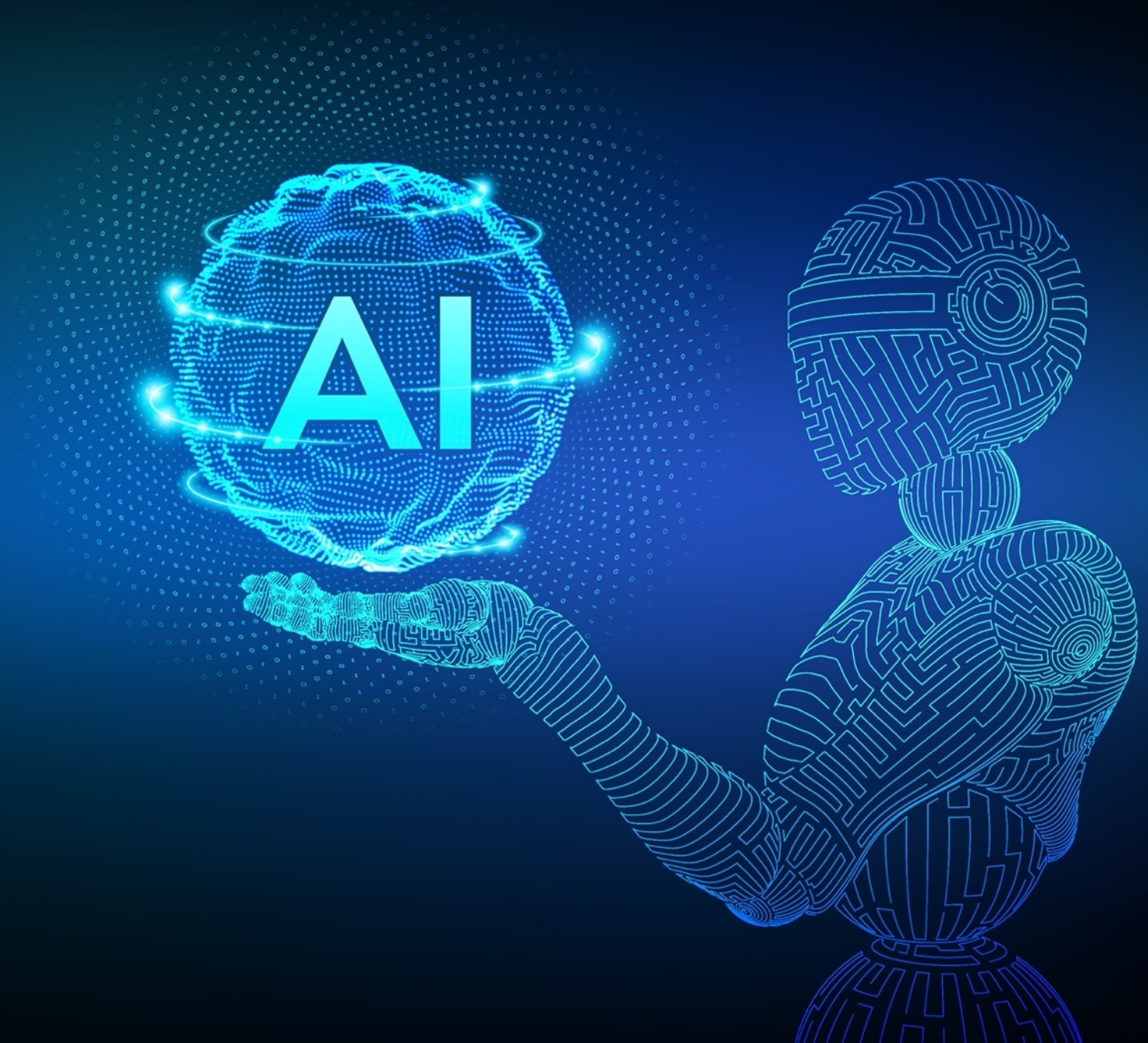
But over the next 12 months, only 35% of executives say their company will focus on improving the governance of AI systems

32%

And only 32% of risk professionals say they’re now involved in the planning and strategy stage of applications of generative AI

Agenda

- The Dual Nature of Generative AI
- Identifying and Mitigating AI-Specific Risks
- Navigating Regulatory Complexities
- Possible Solutions
- What are the risks?
- Collaborative Defense in AI



The Dual Nature Of Generative AI

Potential of Generative AI

- Automating content creation like reports, code, designs
- Natural language capabilities for chatbots, virtual assistants
- Democratizing data analysis and scientific exploration
- Personalizing education and customized recommendations
- Streamlining business workflows and processes
- Enhancing creative outlets like art, music, writing
- Generating insights from proprietary data

Inherent Risks

- Potential misinformation spread or copyright infringement
- Biases and harmful content generation
- Data exposure, poisoning and manipulation
- User profiling and loss of privacy
- Account impersonation, fraud generation
- Deepfakes and unauthorized impersonation
- Theft or unauthorized sharing of sensitive data

Key Areas of Risk

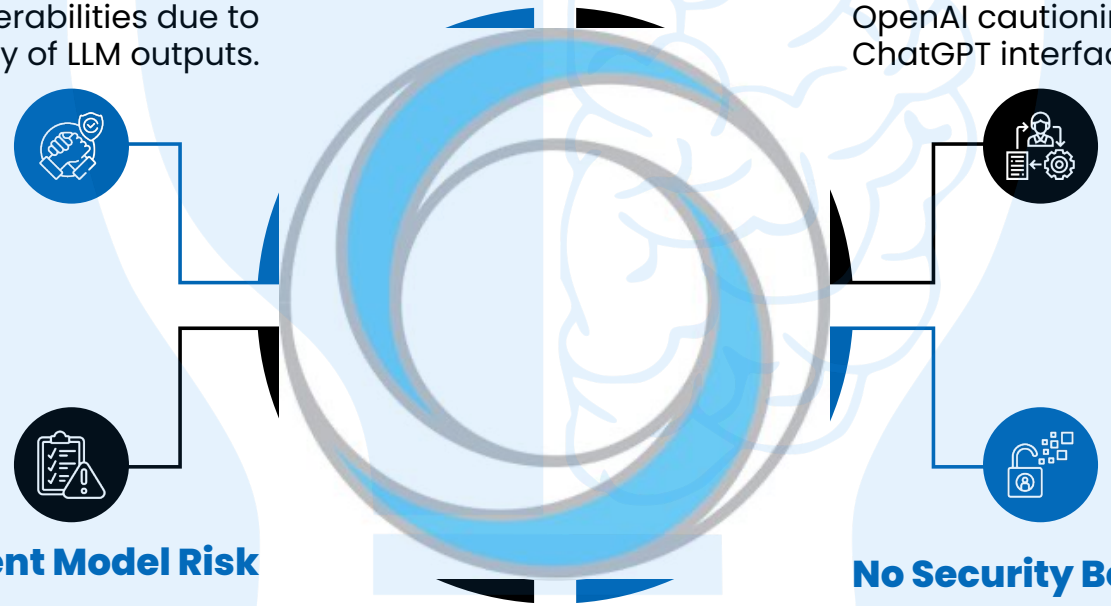
Generative AI Projects Pose Major Cybersecurity Risk To Enterprises

Trust Boundary Risk

Trust boundaries in opensource development ensure security and reliability, but enabling LLMs to access external resources can introduce vulnerabilities due to the unpredictability of LLM outputs.

Data Management Risk

Underlying model risks in LLMs include inadequate AI alignment & overreliance on generated content, with OpenAI cautioning users about these concerns in the ChatGPT interface



Inherent Model Risk

Data-management risks in ML systems can lead to unintentional data leaks and intentional training-data poisoning, compromising the security, effectiveness, and ethics of models like LLMs

No Security Best Practices

Open-source adoption of generative AI can lead to risks like improper error handling and insufficient access controls, allowing attackers to gather sensitive information or exploit vulnerabilities, and enabling users to overstep their permissions

Strategies To Mitigate Generative AI Risks

01. Implement strong data governance around access, lineage, and consent for training data

02. Perform extensive testing to detect biases, distortions, and blind spots prior to deployment

03. Use ethical frameworks and red teaming to surface potential abuses early in the development process

04. Continuously measure how models perform on distorted or adversarial inputs

01. Employ sandboxed development environments and model risk management procedures

02. Monitor models at runtime for signs of data poisoning or model deterioration

03. Rate limit generative models to slow down malicious automation

04. Create triggers to failsafe to safe outputs when models face unpredictable inputs

01. Use differential privacy, federated learning, and other techniques to protect data

02. Build in watermarking techniques to track progeny models and detect theft

03. Employ multiple generative models for diversity & to mask individual model biases

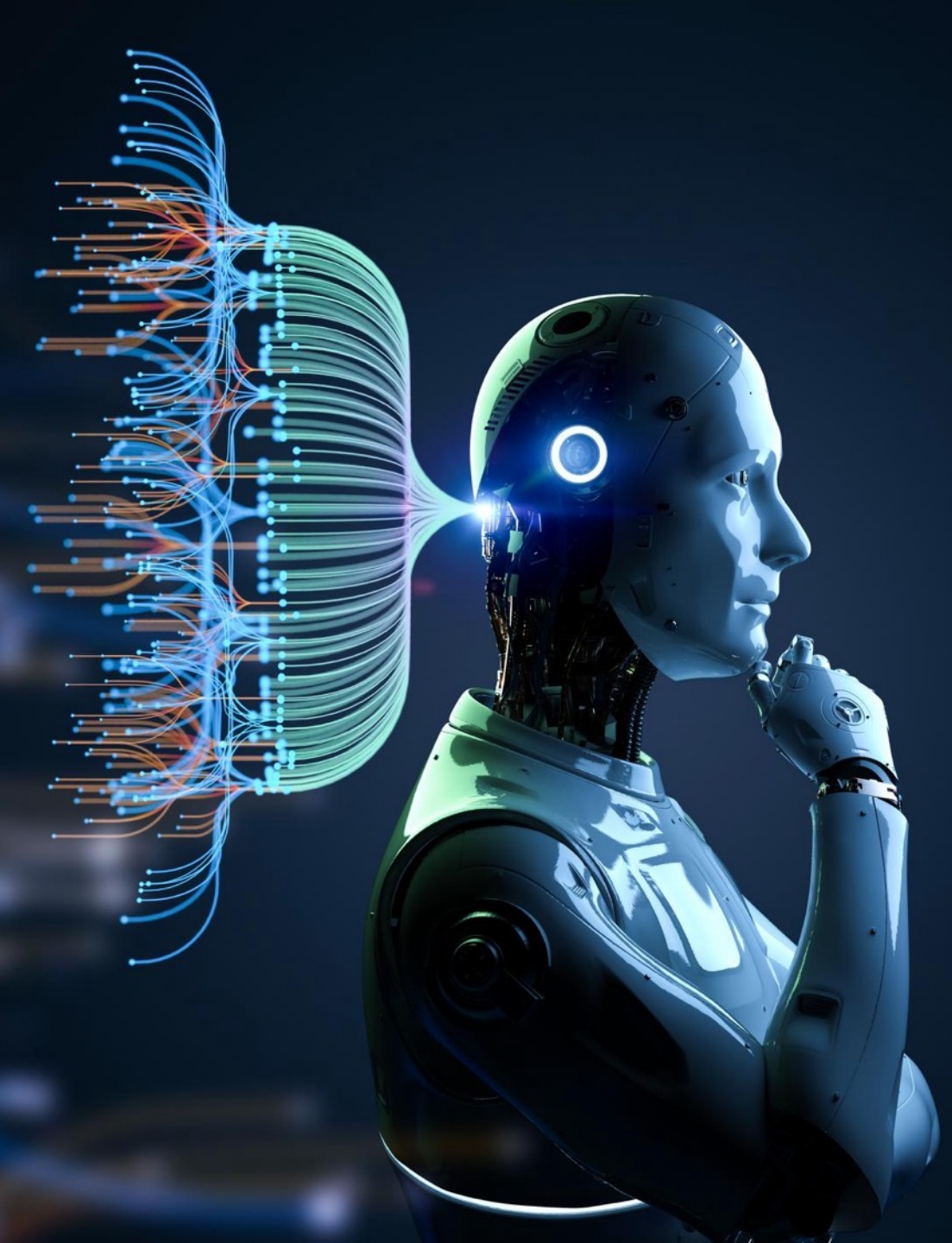
04. Maintain rigorous version control & memorialize model provenance end-to-end

Regulatory Complexities

Current cybersecurity regulations struggle to fully address the emerging data protection, safety, transparency, accountability, and misuse risks introduced by fast-evolving generative AI systems.

Overall, current cybersecurity regulations struggle to fully address the emerging data protection, safety, transparency, accountability, and misuse risks introduced by fast-evolving generative AI systems

- Quality and safety regulations written for standard software struggle to encompass emergent behaviors of complex AI systems. Hard to audit
- Data privacy regulations were not designed to cover synthetic but realistic data generated by AI systems.
- Protecting proprietary AI model IP while also enabling oversight for accountability is a challenging balance
- International variances in AI laws and norms lead to governance gaps that threat actors can exploit



Evolving Compliance Landscape

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

Artificial Intelligence Risk Management Framework (AI RMF 1.0)

AI Frameworks

Established frameworks, such as the NIST AI risk management framework & the ISO framework for AI systems using machine learning, are a good start for designing and deploying trusted AI applications

So too are industry requirements and norms, such as guidance from the Office of the Comptroller of the Currency, Consumer Financial Protection Bureau & the Federal Reserve. And there are more than 800 national AI policies from more than 69 countries, territories & the EU

Evolving Compliance

Global Approach

United States: The Administration and Congress are taking initial steps to produce legislation to regulate AI and using interim measures, such as the White House's recently announced voluntary agreement with seven prominent generative AI companies to provide minimum guardrails for safety, security and public trust, as safeguards. Link

EU and UK: The EU is expected to finalize the EU AI Act, which will classify AI usage based on risk levels, by late 2023, and a white paper issued by the UK government in March empowers sectoral regulators to regulate AI within their jurisdictions and indicates the government's plan to establish central functions to support sectoral regulators. Link

China: In June, China issued its first regulations on generative AI technology, introducing significant obligations for service providers, including content monitoring, marking and data sourcing, while emphasizing the protection of users' personal information through agreements outlining responsibilities. Link

RISK STAKEHOLDERS

01. Chief Information Security Officer (CISO)

02. Chief Data Officer

03. Chief Compliance Officer

04. General Counsel

05. Internal Audit

06. Chief Financial Officer

Chief Information Security Officer (CISO)

Generative AI Threat Landscape

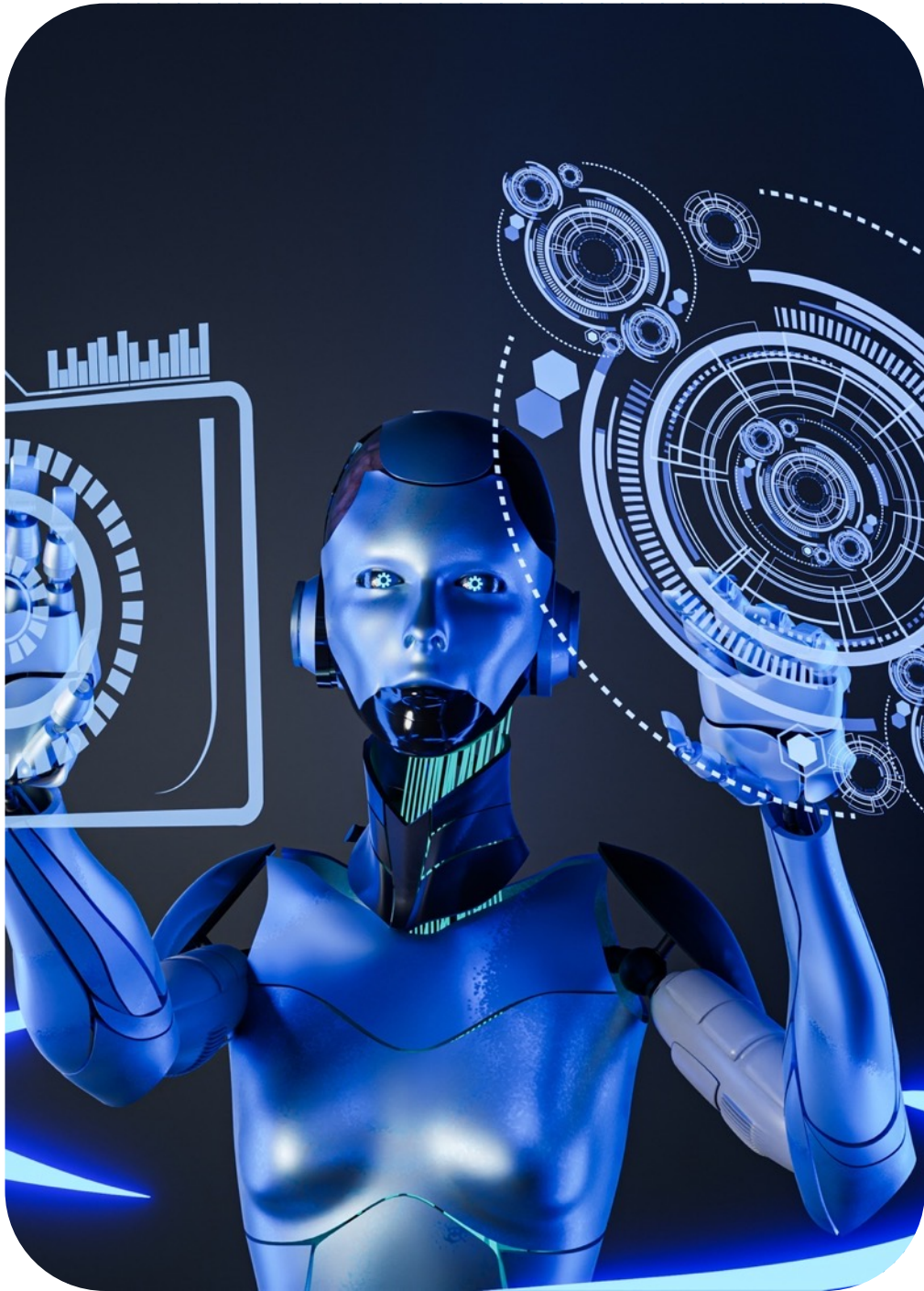
- Sophisticated phishing threats.
- Deep fake video or audio impersonations.
- Offensive AI capabilities mapped to MITRE ATT&CK framework

Cyber Defense Protections

- Protect proprietary language, foundational models, and new content.
- Countermeasures: Automate, update, and upgrade

Action Steps

- Assess access privileges.
- Build EDR platforms using generative AI.
- Evaluate model vulnerabilities.
- Prepare for high-resolution threat models.
- Establish data loss prevention controls.
- Protect internal/local generative AI models



Chief Data Officer

Generative AI Data Risks

- Exacerbation of data and privacy risks
- Unauthorized access, bias, and data loss

Risk-Mitigation Actions

- Enhance data governance protocols.
- Monitor and protect data sharing to external AI models.

Chief Compliance Officer

Regulatory Landscape

- New regulations and stronger enforcement
- Mapping AI applications to existing laws

Risk-Mitigation Actions

- Upgrade regulatory reporting & Monitor FTC actions
- Assess compliance posture
- Update core compliance artifacts
- Establish strong model governance processes

General Counsel

Legal Risks

- IP exposure, secondary data uses, litigation, and investigations

Risk-Mitigation Actions

- Limit IP exposure.
- Guard against improper secondary uses of data
- Plan for litigation and investigations

Chief Financial Officer

Financial Risks

- “Hallucination ” risk on financial facts, errors in reasoning and overreliance on outputs requiring numerical computation.

Risk-Mitigation Actions

- Identify internal controls and statutory requirements.
- Inventory financial tasks.
- Develop HR upskilling and reskilling plan

Internal Audit

Auditing Challenges

- New methodologies, supervision, and skill sets
- Generativity, emergent abilities, lack of grounding, and API accessibility
- It’ s difficult and ineffectual to assess the risks that generative AI systems pose independent of the context in which they are deployed

Risk-Mitigation Actions

- Collaborate with stakeholders
- Adapt risk assessment process
- Audit core data sets
- Design audit plans for AI systems, models, and outputs

Bottom Line

1. Ultimately, the promise of generative AI rests with your people
2. Invest in them to know the limits of using the technology as an assistant, co-pilot, or tutor, even as they exploit & realize its potential
3. Empower your people to apply their experience to critically evaluate the outputs of generative AI models – after building your enterprise risk guardrails
4. Every savvy user can be a steward of trust

Case Study

The Opportunities And Risks Of Building A Generative Ai Powered Medical Consultation Chatbot

They trained the AI model using years of historical patient data including symptoms, diagnoses, and treatments

Title/Position	Actions
Chief Data Officer	✶ Ensured accurate, unbiased training data
Chief Compliance Officer	✶ Verified compliance with healthcare regulations on data use
Chief Privacy Officer	✶ Advocated privacy-by-design approach for user data handling
Chief Technology Officer	✶ Set up dedicated instance separating user data from operations
Legal Department	✶ Negotiated contractual data protections with AI vendor
Chief Information Security Officer	✶ Designated chatbot as high priority crown jewel for security
Internal Audit	✶ Developed risk assessment and audit plan focused on reliability, legal and compliance risks

Key Takeaways

01. Balancing Innovation & Risk

We focused on the transformative potential of generative AI while highlighting the inherent security vulnerabilities it introduces. We learned to strike a balance between innovation and risk mitigation, enabling us to harness the power of AI while safeguarding their digital assets

02. Actionable Strategies for AI Security

We gained practical strategies to identify, assess, and mitigate AI-specific risks. From algorithmic biases to adversarial attacks, with actionable insights to fortify AI systems against emerging threats.

03. Collaborative Defense

We explored how collective defense efforts, including information sharing, can contribute to securing AI ecosystems effectively

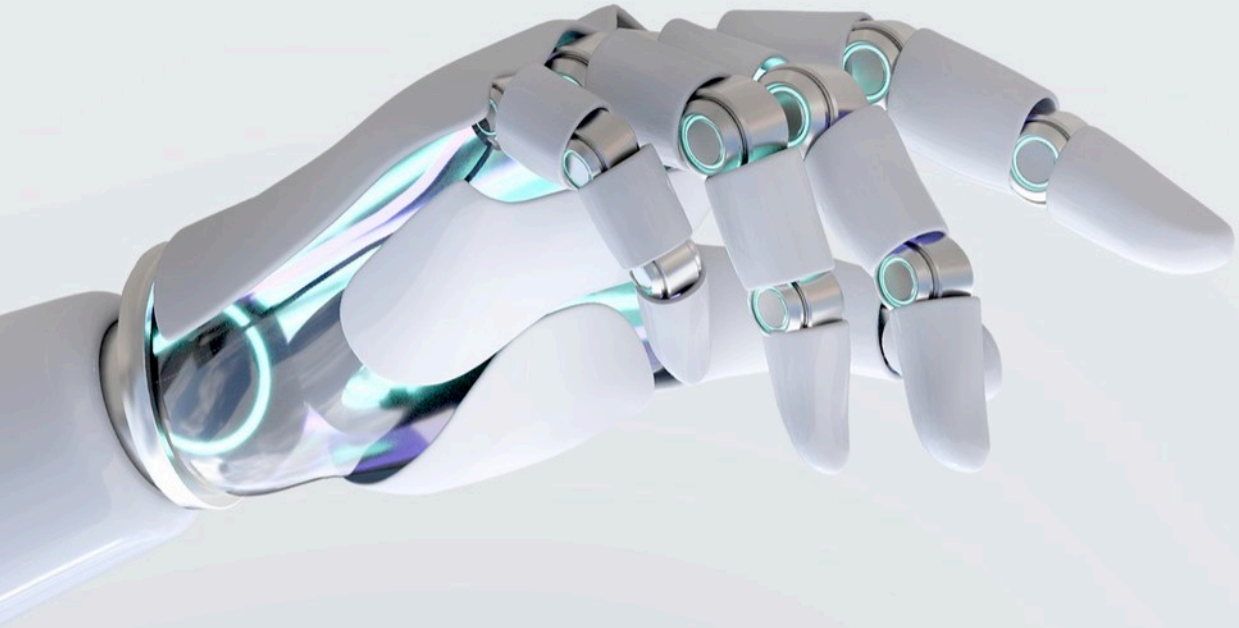
- Cross-functional collaboration on governance, privacy, security, compliance
- Aligned data practices to healthcare regulatory obligations
- Prioritized reliability, auditability, and responsible AI practices

Readings & References

01. https://www.pwc.com/us/en/tech-effect/ai-analytics/responsible-ai-for-generative-ai.html

02. https://www.mayerbrown.com/en/perspectives-events/publications/2023/09/what-boards-need-to-know-regarding-the-forthcoming-artificial-intelligence-related-legal-frameworks-and-what-they-can-do-to-prepare

03. https://www.gartner.com/en/articles/what-it-takes-to-make-ai-safe-and-effective



QUESTIONS?

Viral Trivedi

Viral@aarmorics.com

Viral Trivedi

Cybersecurity Certified CISO - Delivering Cyber Resilience