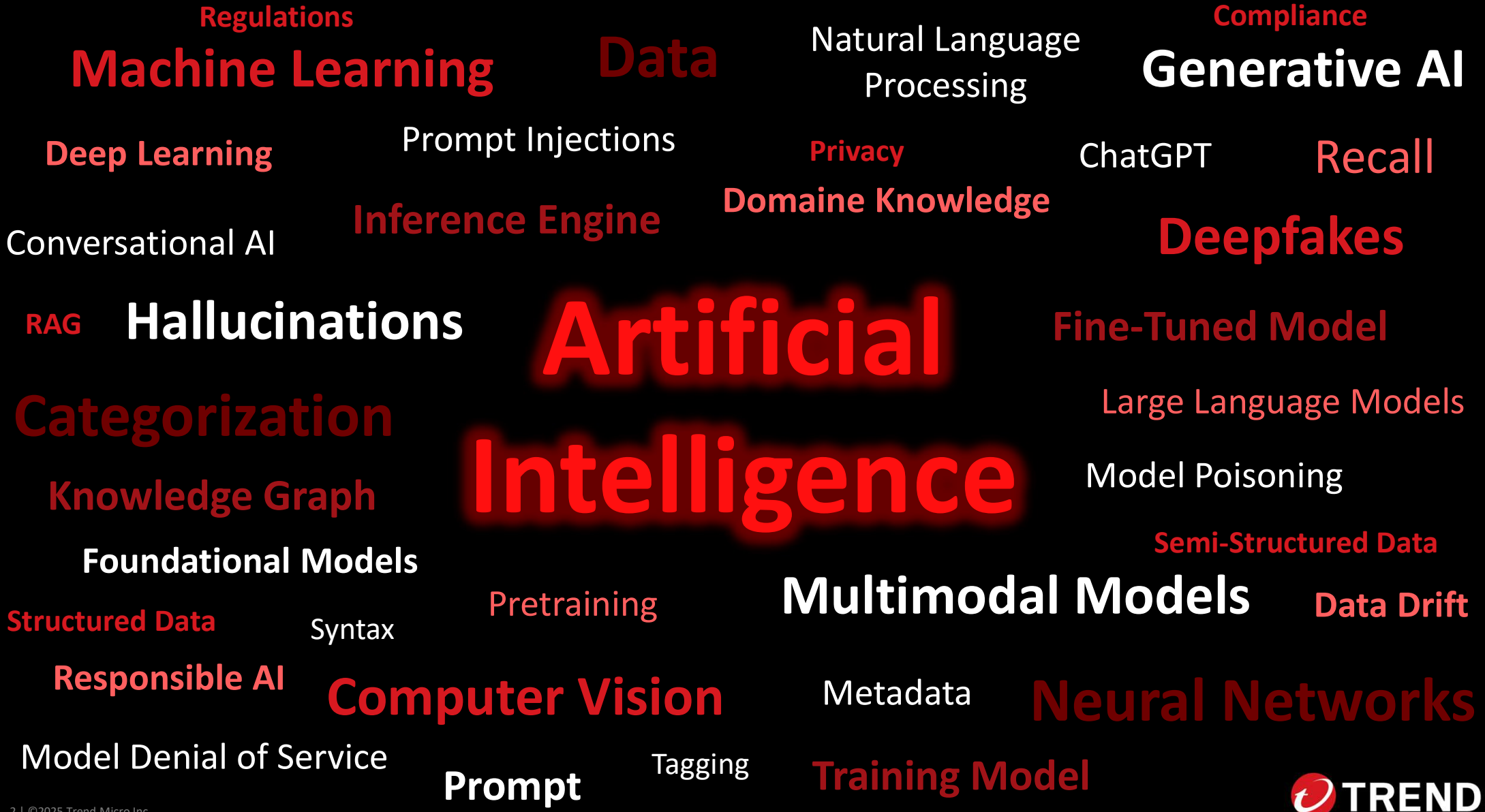




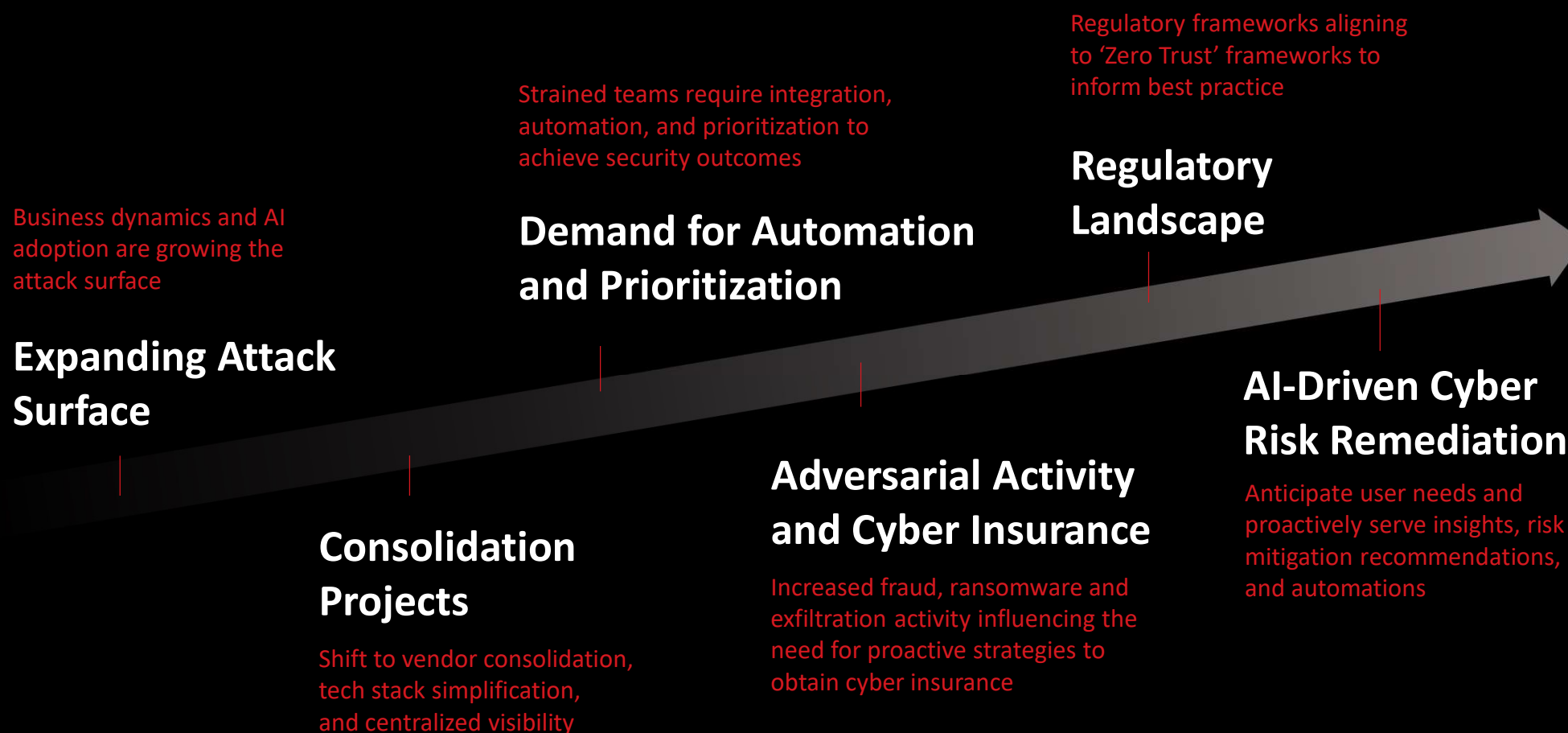
# AI is ruining my life: Group therapy for security leaders

Bharat Mistry





**Feeling  
Overwhelmed?**



J(0      Include Agent here, maybe Agentic Automation and Prioritization

Cyber Insurance seems out of place with the rest

Josiah Hagen (RD-NA); 2024-07-22T21:20:13.016



# Adversarial AI in 2025 and Beyond

Building modern threat models in the AI era



## Managing AI Risk in 2025 and Beyond...

**Rogue AI**

**Scaling Fraud**

**Data & Privacy**

## Snímka 6

---

J(0

Agentic Systems are the enabling technology for Rogue AI. I'd take Agentic Systems out here.

If you need a 3rd risk: Data Loss or Data Privacy violation

Scaling Fraud is now, so is Data Loss

Rogue AI is next year

Attack automation is in the future

Josiah Hagen (RD-NA); 2024-07-22T21:22:44.985



# Rogue Three



## Subverted Rogues

LLM operates against intended use via attacker



## Malicious Rogues

Deployed by attackers to steal computing resources



## Accidental Rogues

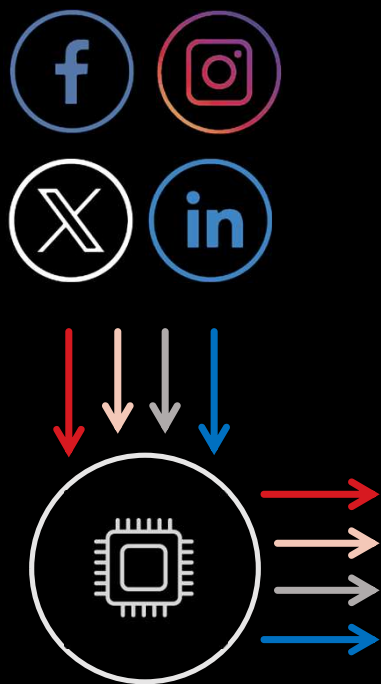
Created by human error or technology limitations

## Snímka 7

---

- 0 Reorder - start w/ subverted, then malicious, and then oh ya.. mistakes are still our biggest problem  
Shannon Murphy (MKT-NA); 2024-07-24T13:25:12.284
- 0 0 Reorder for impact; introducing the new concept about people subverting your AI systems first; and then malware, and then config problems;  
Shannon Murphy (MKT-NA); 2024-07-24T13:25:40.090

# Contending with Fraud Effectiveness and Scale



Hi John,

It was great catching up with some of your co-workers at last week's "Lawyers in Tech" meetup. I saw you weren't able to make it... See [this link](#) for some fun photos from the event!

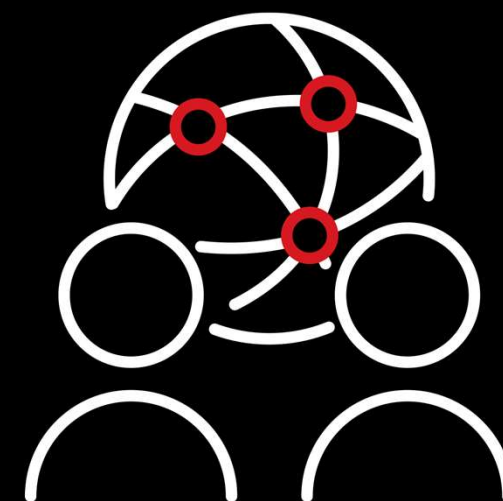


## AI-Driven Phishing Effectiveness

- AI can help improve quality and personalize phishing & fraud emails
- Even greater effectiveness
- Even harder for employees to identify

See: "The Future of Whaling Attacks: AI-Powered Harpoon Whaling", Trend Micro Forward Threat Research, 2023

# Contending with Fraud Effectiveness and Scale



**Mass Misinformation  
Spread via Social Networks**

**Snímka 9**

---

- 0

[@Ashley Savoie (MKT-NA)] these need to be built out a bit more. I updated the phishing related slide on 8; it's a little busy but I think we could consider adding some screenshots here of different misinformation campaigns (i.e. White House on fire, celebrity stuff, election campaign videos)

Shannon Murphy (MKT-NA); 2024-07-26T15:39:33.517
- ASO 0

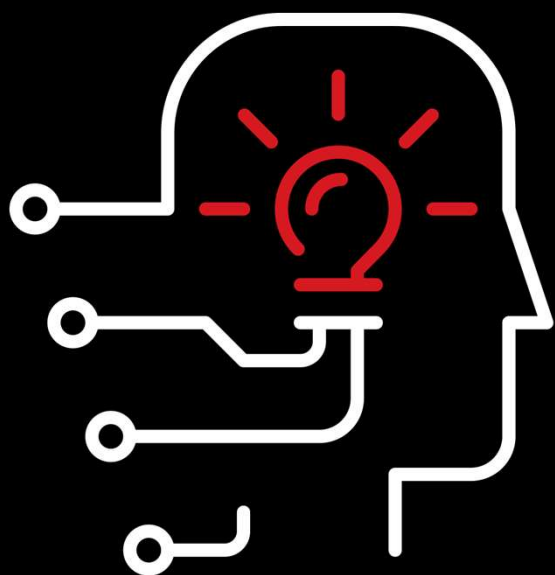
Thoughts?

Ashley Savoie (MKT-NA); 2024-07-26T17:38:23.716
- 0 1

Good

Shannon Murphy (MKT-NA); 2024-07-26T21:53:11.992

# Contending with Fraud Effectiveness and Scale



**Synthetic Media  
Deepfake and Audiofake**

World / Asia

## Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'

By Heather Chen and Kathleen Magramo, CNN  
2 minute read · Published 2:31 AM EST, Sun February 4, 2024

f x e

---

Pro > Security

## Top WPP advertising executive hit by scammers using voice cloning attack

**News** By Sead Fadilpašić published May 13, 2024

AI-powered scams are getting more dangerous

f x p r e

TECH · DEEPFAKES

## A deepfake 'CFO' tricked the British design firm behind the Sydney Opera House in \$25 million scam

BY **PRARTHANA PRAKASH**  
May 17, 2024 at 7:32 AM EDT

e

## Snímka 10

---

ASO

Note from Shannon:

Need to expand on or communicate in the context of an attack workflow or show ChatGPT instance (we did this w/ the previous version of the talk)

Ashley Savoie (MKT-NA); 2024-07-25T11:42:29.315

1

Also need to build out a bit more; maybe provide screenshots of the WPP attack and the UK architectural firm being target of deepfake - not just sci fi; actually coming to fruition;

[@Ashley Savoie (MKT-NA)]

Shannon Murphy (MKT-NA); 2024-07-26T15:40:22.974

# Data Privacy Challenges

## Tracking and Recognising

AI can track and identify people across different devices and places

## Discrimination and Bias

AI can make unfair decisions based on biased data

## Lack of Transparency

Some AI systems are hard to understand or challenge.

## Data Exploitation

People often don't know how much data their devices collect and share.

## Prediction

AI can predict sensitive information from seemingly harmless data





# Securing the AI Transformation

Combining visibility and governance

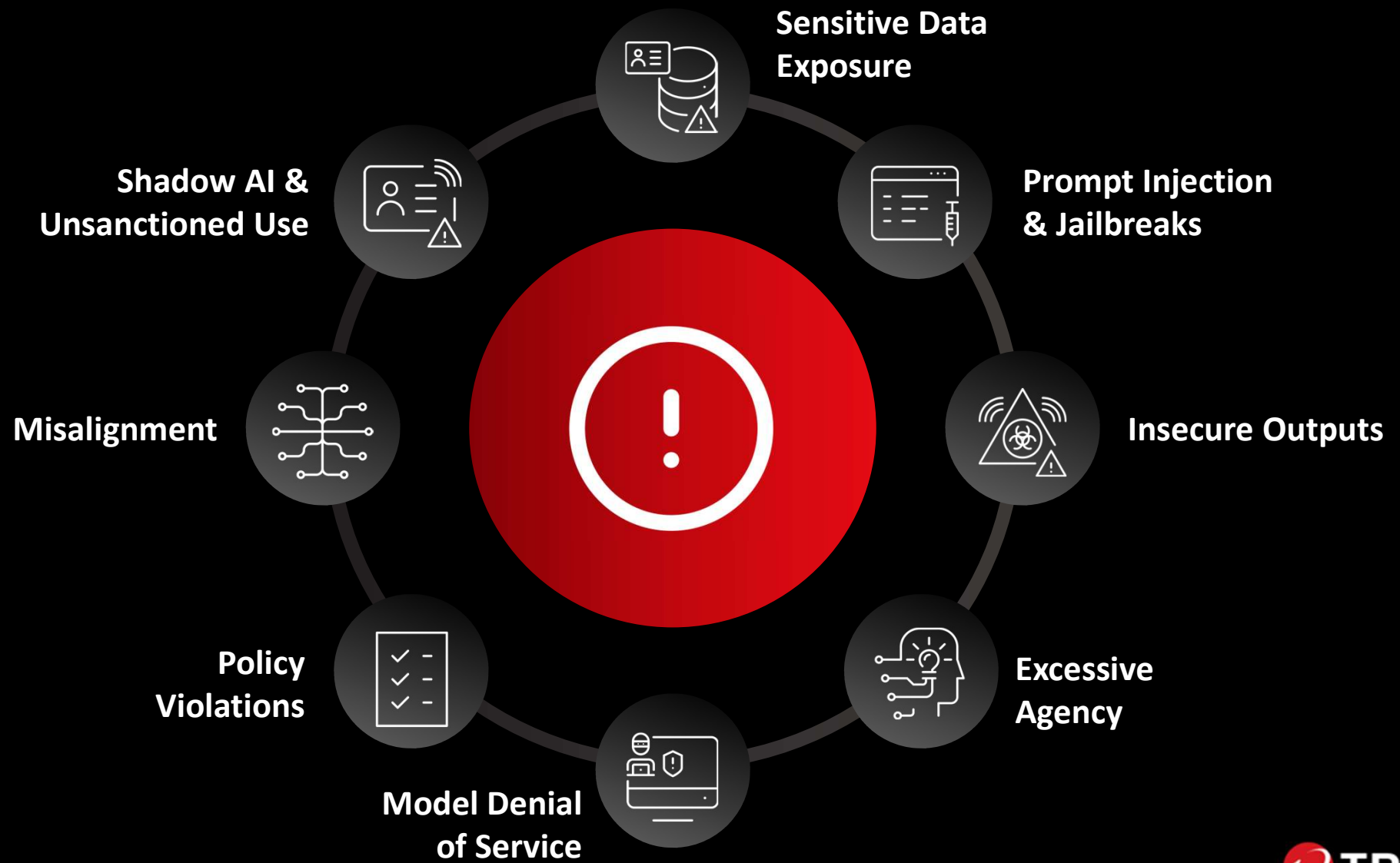




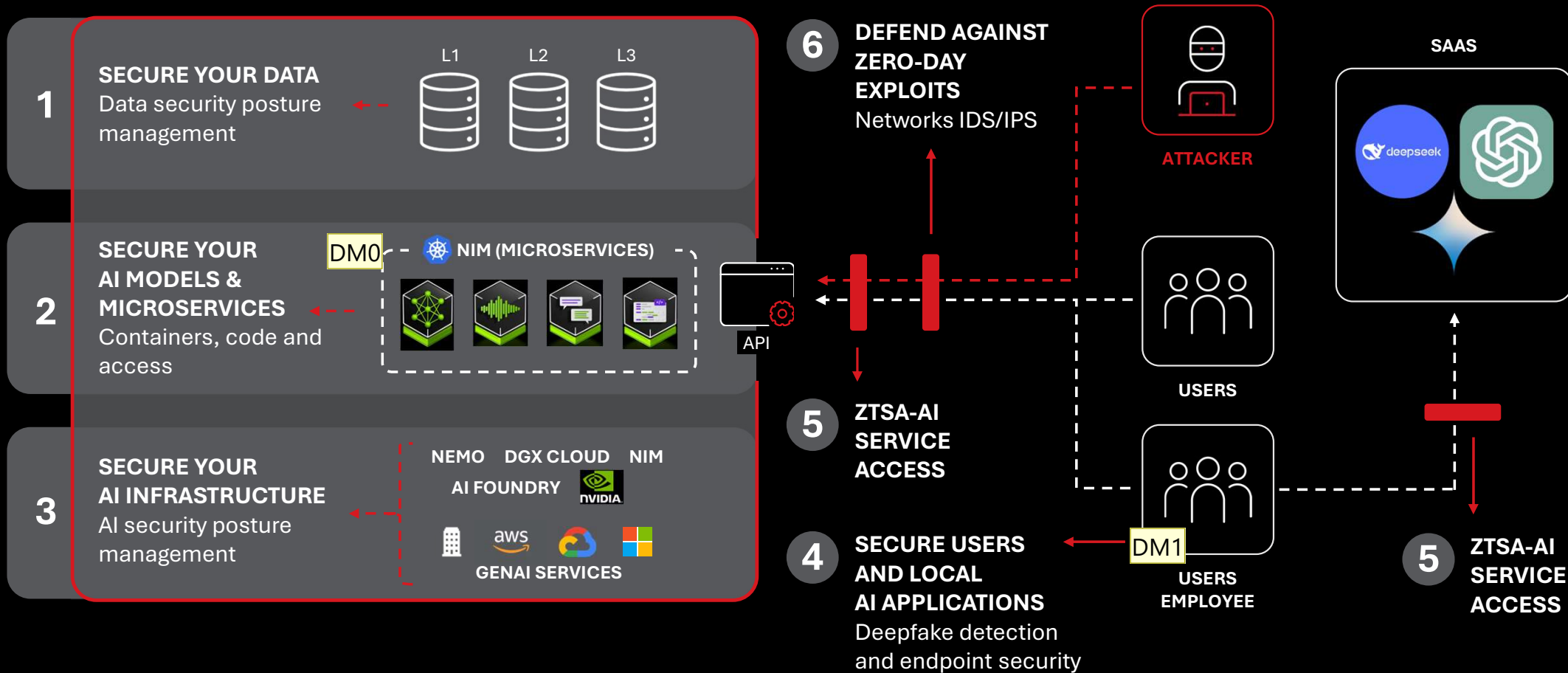


## Executive Team Expectations

**THIS IS YOUR WORK NOW. YOUR  
RESPONSIBILITY.**



# Security For AI Blueprint





## Snímka 15

---

### DM0

So just yesterday, I noticed that for these two elements of our blueprint, they don't line up perfectly well with our Security for AI End-To-End Protection table. (Slide 13).

We do have "Models" listed - but in the table the Trend Vision One solution / capabilities listed are ZTSA AI Secure Access and ZTSA Access Controls and NOT Container Security.

Be good to get Fernando's take on this.

[@Bharat Mistry (PM-EU)]

Dave McDuff (PM-NA); 2025-05-01T12:54:54.889

### DM0 0

If this is a change, please let me (us) know as it needs to be propagated to the many decks we have that use this slide.







Dave McDuff (PM-NA); 2025-05-01T12:58:01.086

### DM1

NOTE this title includes Local AI Apps

Dave McDuff (PM-NA); 2025-05-01T13:00:25.151

# Security for AI: End-to-End Protection

		SECURITY CHALLENGES	SECURITY CONTROLS	TREND VISION ONE
Data		Sensitive information blind spots	Data Security	Data Security posture management
Models		Model poisoning and improper model usage	Implement guardrails for AI API's request/prompt (inbound) & responses (outbound)	ZTSA AI Service Access
Microservices		<b>DM0</b> Vulnerabilities in AI supply chains and microservices architecture	Security validation on CI/CD pipeline and implement container controls	Code security Container security
Infrastructure		Security risks in AI model deployment and resource exhaustion attacks	Infrastructure posture management	AI-SPM API Security AI-DR
Network		Exploiting vulnerabilities in AI infrastructure and hybrid cloud environments	Network Security	Network IDS/IPS TippingPoint
Users		<b>DM1</b> Secure design and mismanagement leading to sensitive data exposure by AI	AI application access control and protect local AI application configurations	ZTSA AI Service Access V1ES - Deepfake detection AI app guard

**DM0** NOTE: There is only a minor mention of Microservices in the Blueprint slide (slide 12) with AI Models? Any updates required here?

[@Bharat Mistry (PM-EU)]

Dave McDuff (PM-NA); 2025-05-01T12:55:53.682

**DM0 0** If this is a change, please let me (us) know as it needs to be propagated to the many decks we have that use this slide.

Dave McDuff (PM-NA); 2025-05-01T12:57:54.884

**DM1** Should this be re-labelled Users and AI Apps to be consistent with the Blueprint slide 12 point number 4?





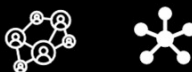

[@Bharat Mistry (PM-EU)]

Dave McDuff (PM-NA); 2025-05-01T13:01:02.605

**DM1 0** If this is a change, please let me (us) know as it needs to be propagated to the many decks we have that use this slide.

Dave McDuff (PM-NA); 2025-05-01T13:01:13.850

# Security for AI: End-to-End Protection

		SECURITY CHALLENGES	SECURITY CONTROLS
Data		Sensitive information blind spots	Data Security
Models		Model poisoning and improper model usage	Implement guardrails for AI API's request/prompt (inbound) & responses (outbound)
Microservices		Vulnerabilities in AI supply chains and microservices architecture	Security validation on CI/CD pipeline and implement container controls
Infrastructure		Security risks in AI model deployment and resource exhaustion attacks	Infrastructure posture management
Network		Exploiting vulnerabilities in AI infrastructure and hybrid cloud environments	Network Security
Users		Insecure design and mismanagement leading to sensitive data exposure by AI	AI application access control and protect local AI application configurations



**Trend Vision One**  
Enterprise Cybersecurity Platform



# Security for AI Blueprint

- What is an LLM application?
- How to integrate security into AI applications architecture?
- The AI Attack Surface in Action
- Threat Modelling for LLM
- Blueprint For Securing AI
- LEARN Architecture Overview

A graphic for the 'Security for AI Blueprint' report. It features a dark background with a stylized profile of a human head on the right, composed of blue and red smoke or energy. A glowing red circular light is visible near the ear area. On the left, the text 'Trend Research' is at the top, followed by 'Unlock safe AI innovation with Trend Micro's'. Below this, 'SECURITY FOR AI BLUEPRINT' is written in large, bold, white letters. Underneath, it says 'for your Datacenter and Cloud'. At the bottom left, there is a QR code and the text 'Ready to transform your AI strategy? Download the BLUEPRINT!'. The Trend Micro logo is in the top right corner.

Trend Research

Unlock safe AI innovation with Trend Micro's

**SECURITY FOR AI BLUEPRINT**

for your Datacenter and Cloud

Ready to transform your AI strategy? Download the **BLUEPRINT!**





**Bharat Mistry**

[Bharat\\_mistry@trendmicro.com](mailto:Bharat_mistry@trendmicro.com)