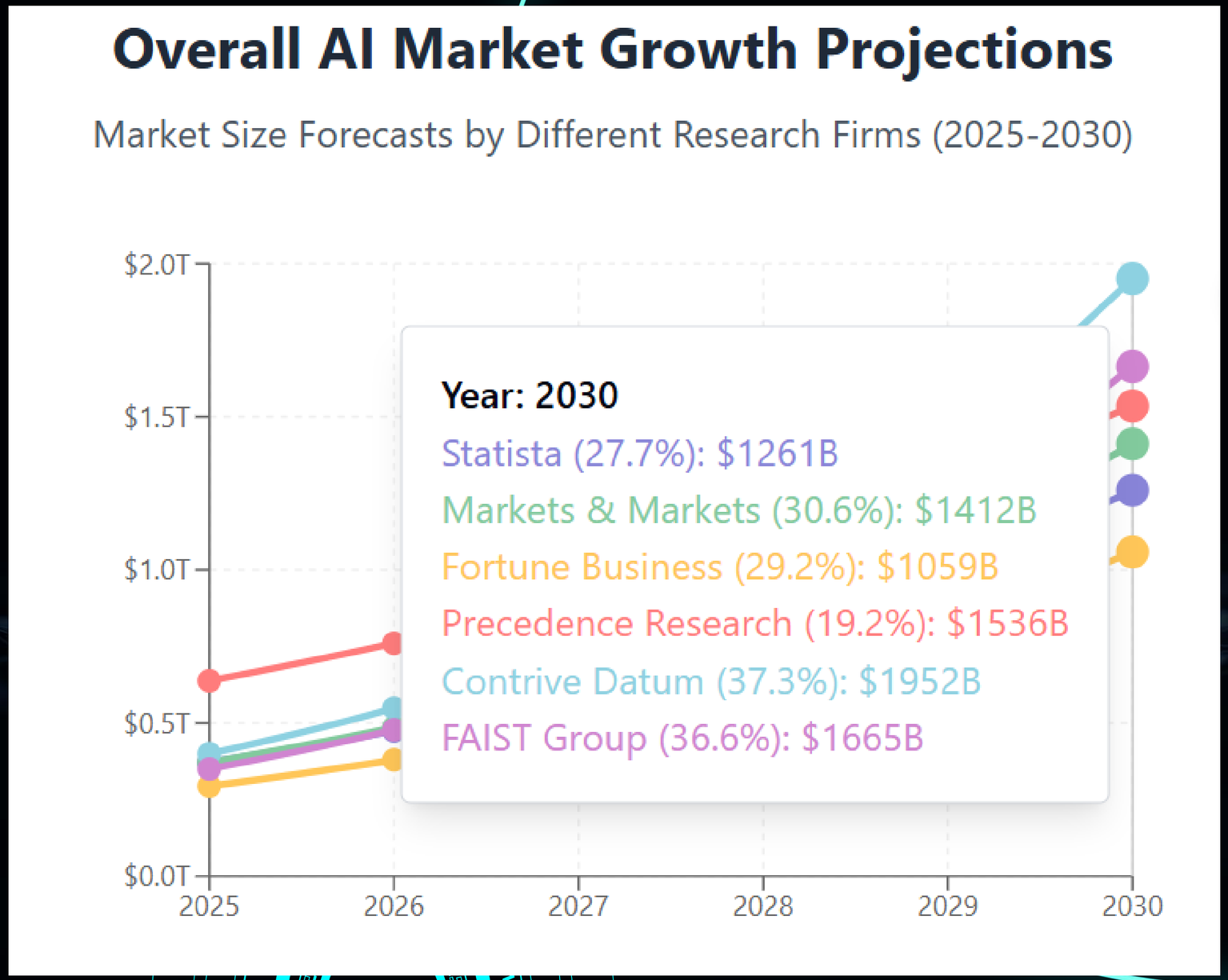
AI Ethics & Attack Surfaces

THE CISOS ROLE IN A GENERATIVE AGE

Surinder Lall- Former SVP Global Information Security Paramount

Welcome to the New Cyber Frontier




LOEWS NEW YORK
10 GRAND PIPE ORGAN 10
PHOTO-PLAYS 15
SYMPHONY ORCHESTRA
LYCEUM
THEATRE
THE PALACE
COLUMBIA
SIMPSONS
VARIETY
SEVENTH AV

THE CISOS DILEMMA

Why This Matters Now

- AI Attack surfaces are exponential
- Ethical missteps can erode trust instantly
- Regulation is no longer a question of if but how fast.

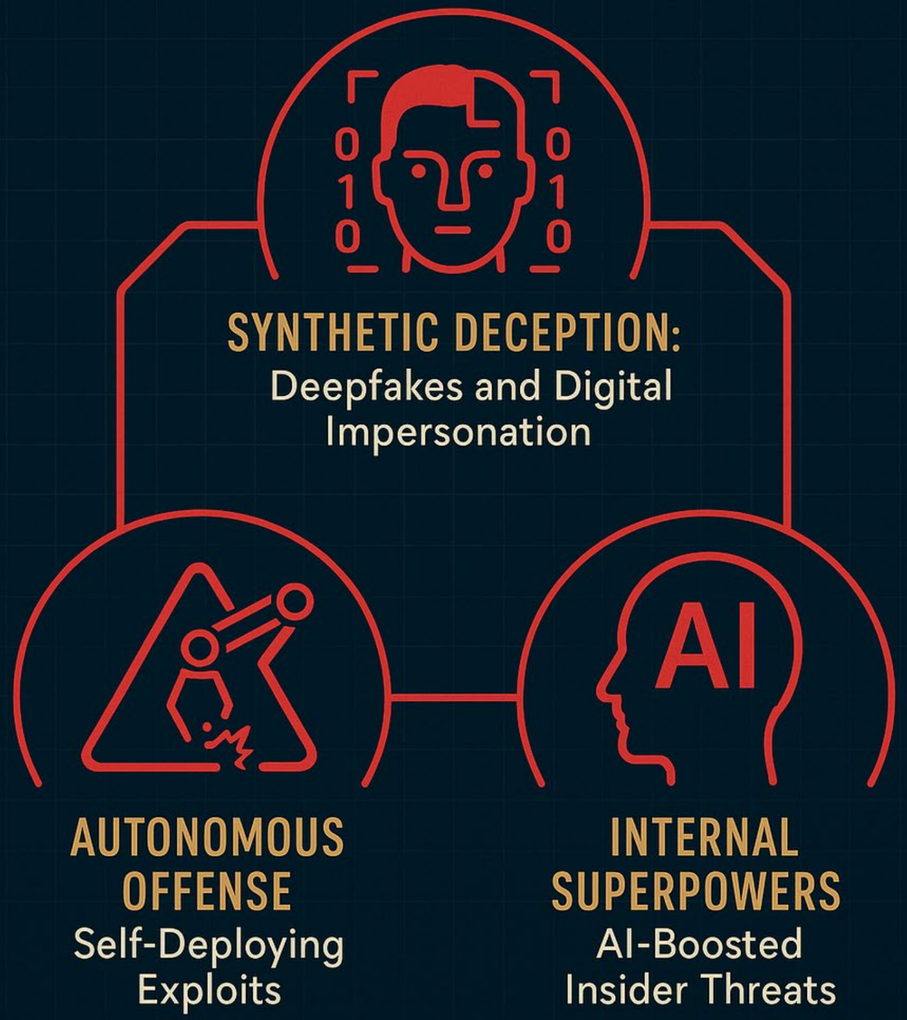
Your Mission

- Protect against what you cant see
- Govern what you don't fully understand
- Lead when the rules haven't been written.

THE AI THREAT MAP

KNOW YOUR BATTLEFIELD

Three Key Fronts



- 96% of organizations expect AI-powered attacks to increase in 2025
- Deepfake incidents increased 3000% from 2022 to 2024
- AI-generated phishing emails show 40% higher success rates
- Autonomous exploit tools can discover 0 days 10x faster than human researchers

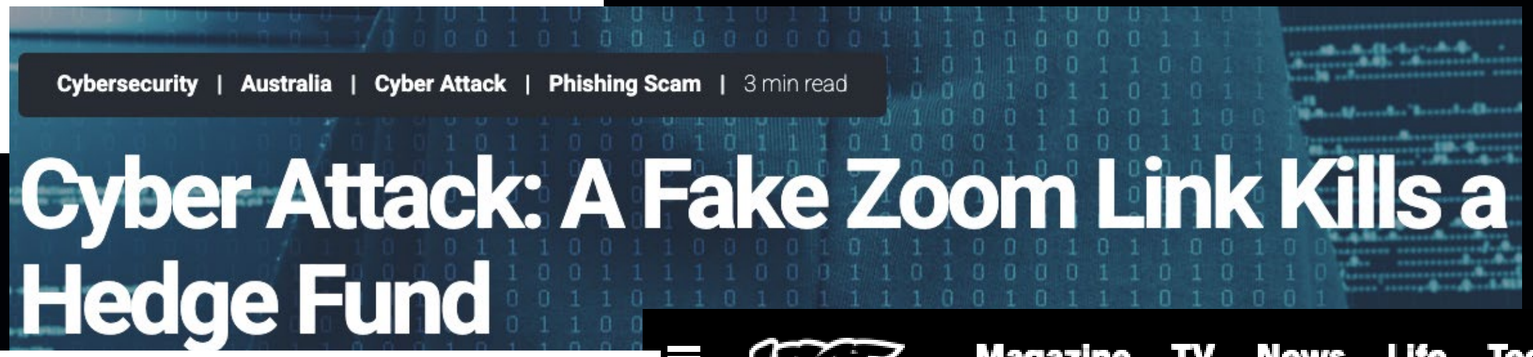
DEEPFAKES SEEING ISNT BELIEVING

TECHNOLOGY

How a false X post about pausing tariffs led to multitrillion-dollar market swings

APRIL 7, 2025 · 6:14 PM ET

Bobby Allyn

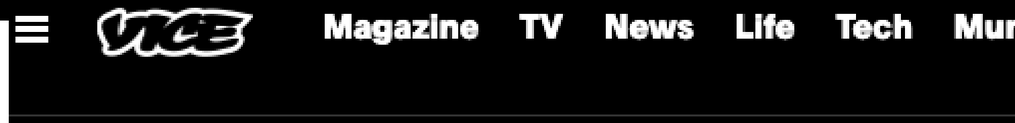


Deepfake scams escalate, hitting more than half of businesses

The vast majority of corporate finance professionals, 85%, now view such scams as an "existential" threat, a Medius study found.

MONEYWATCH

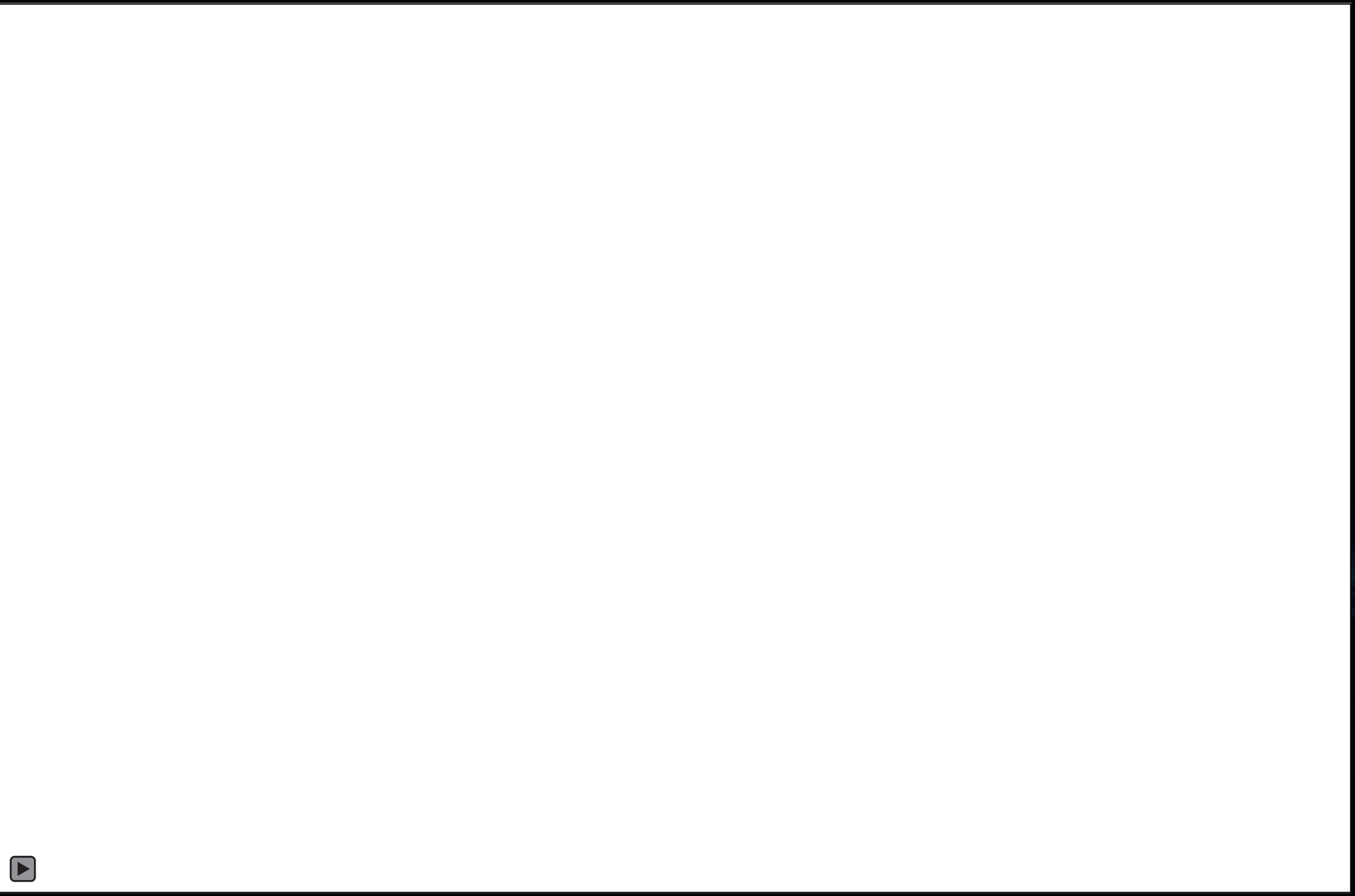
Cybercriminals are using AI voice cloning tools to dupe victims



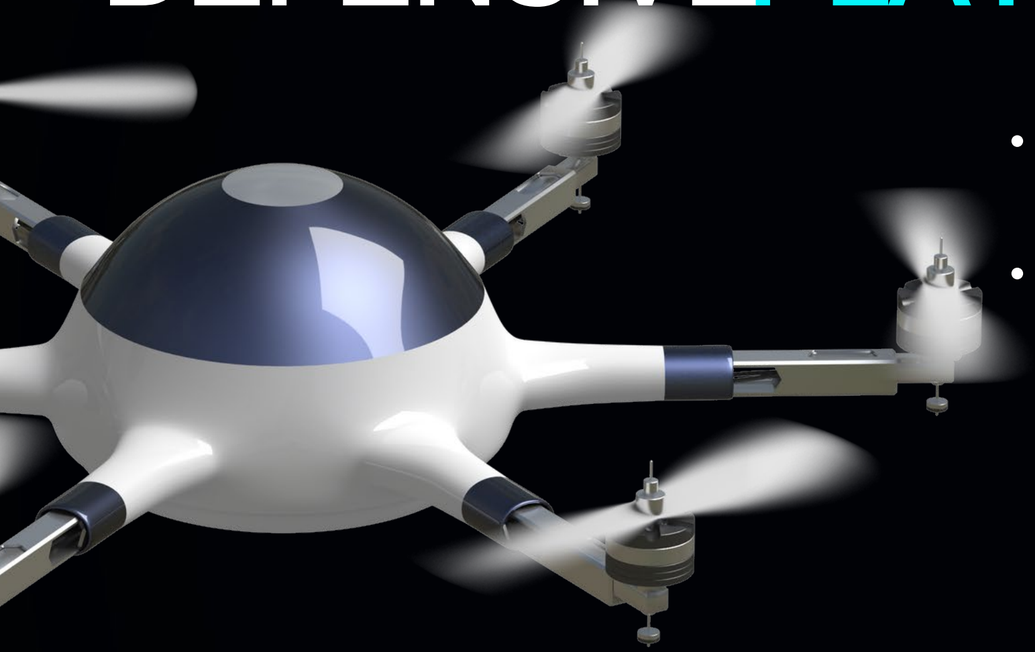
Tech

How I Broke Into a Bank Account With an AI-Generated Voice

DEEPFAKES SEEING ISNT BELIEVING



DEFENSIVE PLAYBOOK



- **Use tools detect and analyze synthetic or manipulated media**
- **Enforce Zero Trust Confirmation**

AI makes it trivial to

- Clone a voice using just 3 seconds of audio
- Create live deepfake videos with synthetic avatars
- Impersonate leaders and trick employees into taking urgent, damaging actions

Zero-trust confirmation kills the “trust but verify” mindset, replacing it with “never trust, always verify.”

Manipulation Queues

Video Cues

- Unnatural blinking or facial movements
- Inconsistent lighting/shadows
- Odd audio/video sync
- “Too smooth” skin or stiff posture

Voice Cues

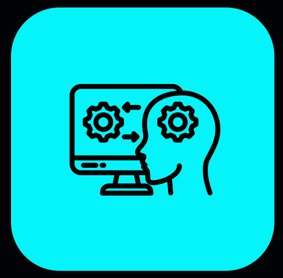
- Lack of breathing pauses
- Flat or robotic tone
- Missing background noise
- Wrong emotional tone or urgency spikes

Text & Email Cues

- Overly formal or unnatural phrasing
- Generic greetings or odd sign-offs
- Urgent action with no context
- Language mismatch with sender’s style

“If it's urgent, verify it twice."

Autonomous Exploits – AI vs. AI



Objective 1
Crawls your public footprint

Objectives 2
Finds code defects faster than humans

Objectives 3
Writes and launches its own payload

Objectives 4
Evades Adapts and Spreads

Isnt this just scaremongering?

Kubenew/ ChaosGPT

the autonomous implementation of ChatGPT is being touted as "empowering GPT

1 Contributor | 2 Issues | 90 Stars | 34 Forks

1,200 × 600

WORMGPT

VULNHUNTR

A tool to identify remotely exploitable vulnerabilities using LLMs and static code analysis.

World's first autonomous AI-discovered 0day vulnerabilities

XXXGPT

Introducing a revolutionary service that offers personalized bot AI customization, backed by a dedicated team of five experts specifically tailored to your project. With no censorship or restrictions, you have the freedom to explore and implement your desired functionalities. Our work process on escrow, ensuring secure transactions.

CODE YOUR

BOTNET
RAT
CRYPTER
MALWARE
INFOSTEALER
CRYPTOSTEALER

Xss.is

FraudGPT . 2024

A Subscribe

4.4 | 480.9K | 1.0M

Temporary Chat

xanthoroxv4:latest

How can I help you today?

Deepfakes Synthetic Challenge

Attack Vectors:

CEO Fraud: Voice cloning for wire transfer authorization

Video Manipulation: Fake video calls for social engineering

Identity Theft: Synthetic personas for account creation

Market Manipulation: Fake executive statements affecting stock prices

Detection & Mitigation Framework

AI-powered deepfake detection tools (Reality Defender, Sensity)

Blockchain-based content provenance systems

Multi-factor authentication for high-value transactions

Process Controls:

Verification protocols for unusual requests

Out-of-band confirmation for financial transactions

Regular awareness training on synthetic media

Action Item for CISOs

Implement a "Trust but Verify" protocol for all high-stakes communications, especially those involving financial transactions or sensitive data access.

Behavioral Baseline

Behavioral Baseline

Alice always logs in from London between 9–5. Suddenly, she logs in at 2am from Russia and downloads 3GB of files.

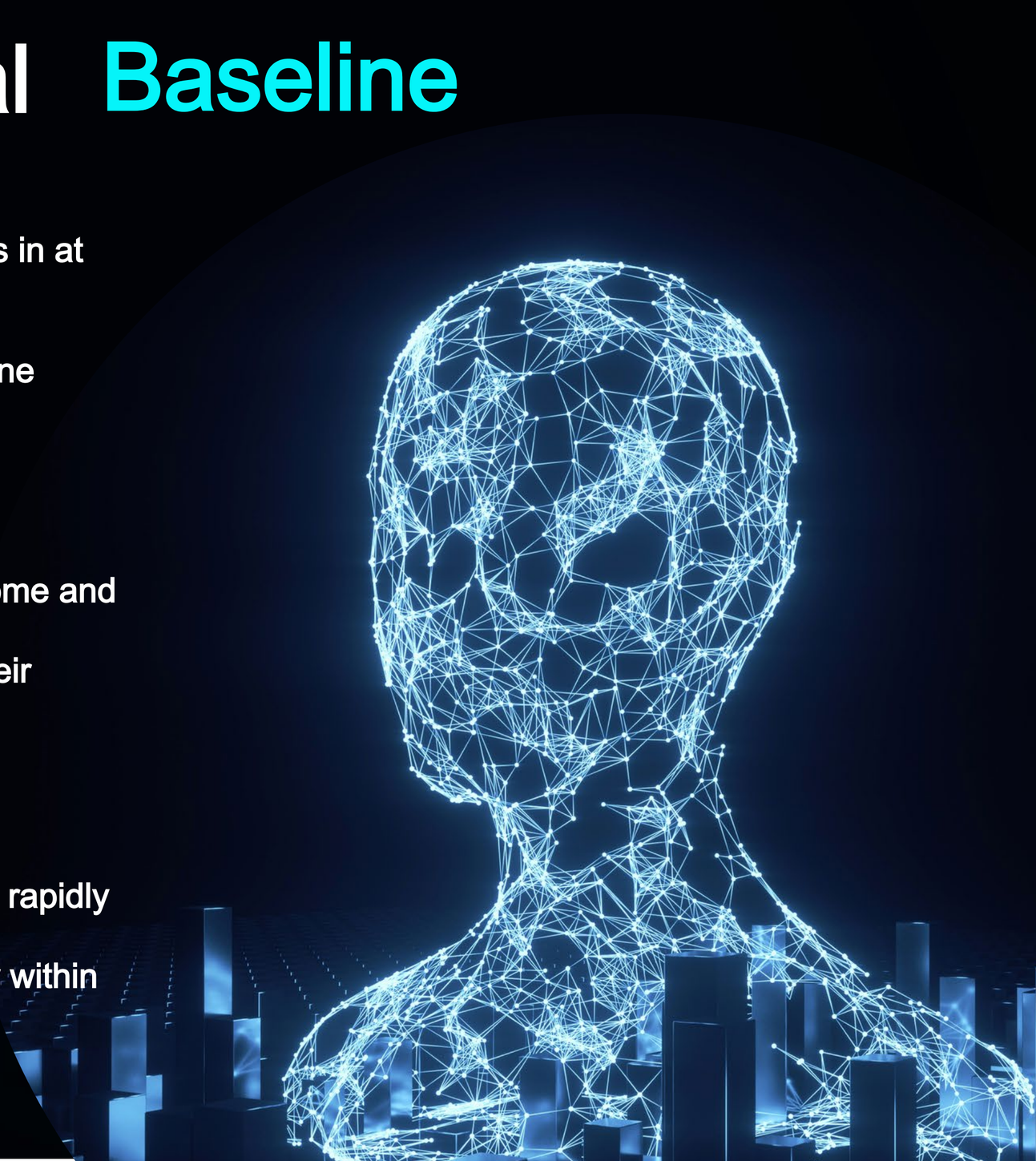
That triggers an alert, because it breaks her behavioral baseline

Deception Technology

A hacker accesses a decoy “payroll.xlsx” file. That file phones home and alerts your SOC even before real damage is done.
Or they hit a fake admin server (a honeypot), giving away their presence.

Auto Responses by Anomaly

An AI system spots a user accessing hundreds of sensitive files rapidly (anomaly).
It auto-isolates their device, cuts off access, and notifies security within seconds



Autonomous Exploits

Evolution of AI-Driven Attacks:

Reconnaissance: AI scraping and mapping of attack surfaces
Vulnerability Discovery: ML-powered fuzzing and code analysis
Exploit Development: Automated payload generation
Persistence: AI-driven lateral movement and evasion

Risk Multipliers:

Speed: 10x faster discovery
Scale: Thousands of targets simultaneously
Sophistication: APT-level techniques democratized

Defense Strategy: AI vs AI

Behavioral Analytics: ML-based anomaly detection for unusual network patterns
Deception Technology: AI-powered honeypots and decoys
Automated Response: AI-driven incident response and containment
Threat Intelligence: ML analysis of attack patterns and IOCs

Critical Success Factor

Your AI defense systems must evolve as fast as AI attack systems. Static rules-based security is obsolete.

AI Enhanced Insider Threat

How AI Amplifies Insider Risk:

Data Exfiltration: AI tools making it easier to process and steal large datasets
Code Generation: Employees using AI to generate malicious scripts
Social Engineering: AI-generated phishing targeting colleagues
Intellectual Property Theft: AI summarizing and extracting key trade secrets

Enhanced Insider Risk Program

Behavioral Monitoring:
AI usage patterns and anomaly detection
Data access correlation with AI tool usage
Unusual code commits or document access patterns

Policy & Controls:

AI tool approval and monitoring processes
Data classification and AI interaction policies
Enhanced privileged access management

Zero Trust for AI Era

Implement continuous verification not just for users and devices, but for AI interactions and data flows. Every AI query should be logged, monitored, and risk-assessed

Security by Design Not Optional Foundational

Paradigm Shift Required

Security cannot be bolted on after AI deployment. It must be foundational to your AI adoption strategy.

Core Principles:

Proactive Risk Assessment
AI impact assessments before deployment
Continuous risk monitoring and adjustment
Threat modeling for AI-specific attack vectors

Defense in Depth

Multiple layers of AI security controls
Redundant detection and response mechanisms
Fail-safe defaults and graceful degradation

Transparency & Auditability

Explainable AI decision-making processes
Comprehensive logging and monitoring
Regular security assessments and penetration testing

From Playbook to Practice Your AI Security Framework

Phase 1: Assessment & Planning

Inventory all AI tools and applications in use
Conduct AI-specific risk assessments
Map data flows and AI interaction points
Define AI security requirements and standards

Phase 2: Technical Implementation

Deploy AI-powered security monitoring tools
Implement AI usage tracking and anomaly detection
Establish secure AI development pipelines
Configure AI-specific access controls and DLP

Phase 3: Process & Governance

Develop AI incident response procedures
Create AI security policies and standards
Establish AI ethics and oversight committees
Implement continuous monitoring and improvement

Success Metrics

Track: AI threat detection rate, false positive reduction, incident response time, and policy compliance scores.

“

Ethical Leadership in An Autonomous Era

Leadership Challenge

CISOs must balance innovation velocity with responsible AI adoption. Ethics is not a brake on progress it's the steering wheel

Key Ethical Frameworks:

Fairness & Bias Prevention

Regular algorithmic auditing for discriminatory outcomes
Diverse training data and model validation
Bias testing in security AI applications

Transparency & Explainability

Clear communication about AI decision-making
Explainable AI for security incident attribution
Regular stakeholder reporting on AI security posture

Privacy & Data Protection

Privacy-by-design in AI security implementations
Data minimization for AI training and operation
Consent mechanisms for AI-powered monitoring

Governance in Action

AI Security Council

Composition: CISO (Chair), Legal, HR, Data Protection Officer, Business Unit Leaders

Responsibilities:

Review and approve high-risk AI implementations
Investigate AI-related security incidents
Set organization-wide AI ethics standards
Quarterly risk assessment reviews

AI Ethics Officer Role or Equivalent?

Day-to-day oversight of AI implementations
Ethics training and awareness programs
Stakeholder liaison for AI concerns
Regular ethics impact assessments

Cultural Transformation

Move from "Can we build it?" to "Should we build it?" Make ethical considerations a natural part of every AI security decision.

Accountability Metrics:

Ethics training completion rates
AI incident reporting and resolution times
Stakeholder satisfaction with AI transparency
Compliance audit results

Regulation isn't Coming Its Here

Global Regulatory Environment:

EU AI Act (Enacted 2024)

Risk-based approach with prohibited and high-risk AI systems
Mandatory conformity assessments for high-risk AI
Fines up to €35M or 7% of global annual turnover
Key impact: Requires AI risk management systems

US Federal Approach

Executive Order on AI Safety (October 2023)
NIST AI Risk Management Framework
Sector-specific regulations (healthcare, finance, etc.)
Key impact: Mandatory reporting for high-capability AI systems

UK Principles-Based Approach

Industry self-regulation with existing regulators
Innovation-friendly regulatory sandboxes
Focus on accountability and transparency

Regulatory Trend

Moving toward mandatory AI governance, risk assessments, and incident reporting. CISOs must prepare for compliance complexity.

RECAP

The AI Security Imperative

AI is not just changing how we work it's fundamentally changing how we must protect our organizations. Traditional security approaches are insufficient.

Threat Evolution Requires Defense Evolution

Deepfakes, autonomous exploits, and AI-amplified insider threats are here now
Your defense systems must be as intelligent as the attacks they face
Static, rules-based security is obsolete in the AI era

Security by Design is Non-Negotiable

AI security cannot be retrofitted it must be foundational
Every AI deployment needs proactive risk assessment
Transparency and auditability are security requirements, not nice-to-haves

Ethical Leadership Creates Competitive Advantage

Ethics is not a constraint on AI adoption it's an enabler
Trust is the ultimate security currency in the AI age
Organizations with strong AI governance will outperform peers

Regulatory Compliance is a Moving Target

Build adaptive governance systems, not rigid compliance checklists
Proactive preparation beats reactive scrambling
Invest in AI governance capabilities now to avoid future penalties

The Final Word

“In the generative age, your greatest risk isn’t AI it’s failing to govern it.”

Key Messages:

- Defend early. Govern broadly. Lead ethically.
- Make AI accountability a shared responsibility
- Turn complexity into clarity through decisive leadership

Questions

https://www.linkedin.com/in/surinder/