Enterprise Risks of Generative AI Services

for the Board of Directors and Executive Leadership

Scott Foote, Phenomenati Consulting

Generative AI is *Everywhere*

Images

Image Synthesis

Deep Fakes

Video

Video Synthesis

Source Code Generation

(e.g., GitHub Copilot)

Creative Design

Sentiment Analysis

Financial Modeling

Collaboration

(e.g., Slack, SalesForce w/ Einstein GPT)

Content Generation

(e.g., Adobe Firefly,

Microsoft 365 Copilot)

Data “Augmentation”

Virtual Assistants and Chatbots

Search

(e.g., Microsoft Bing)

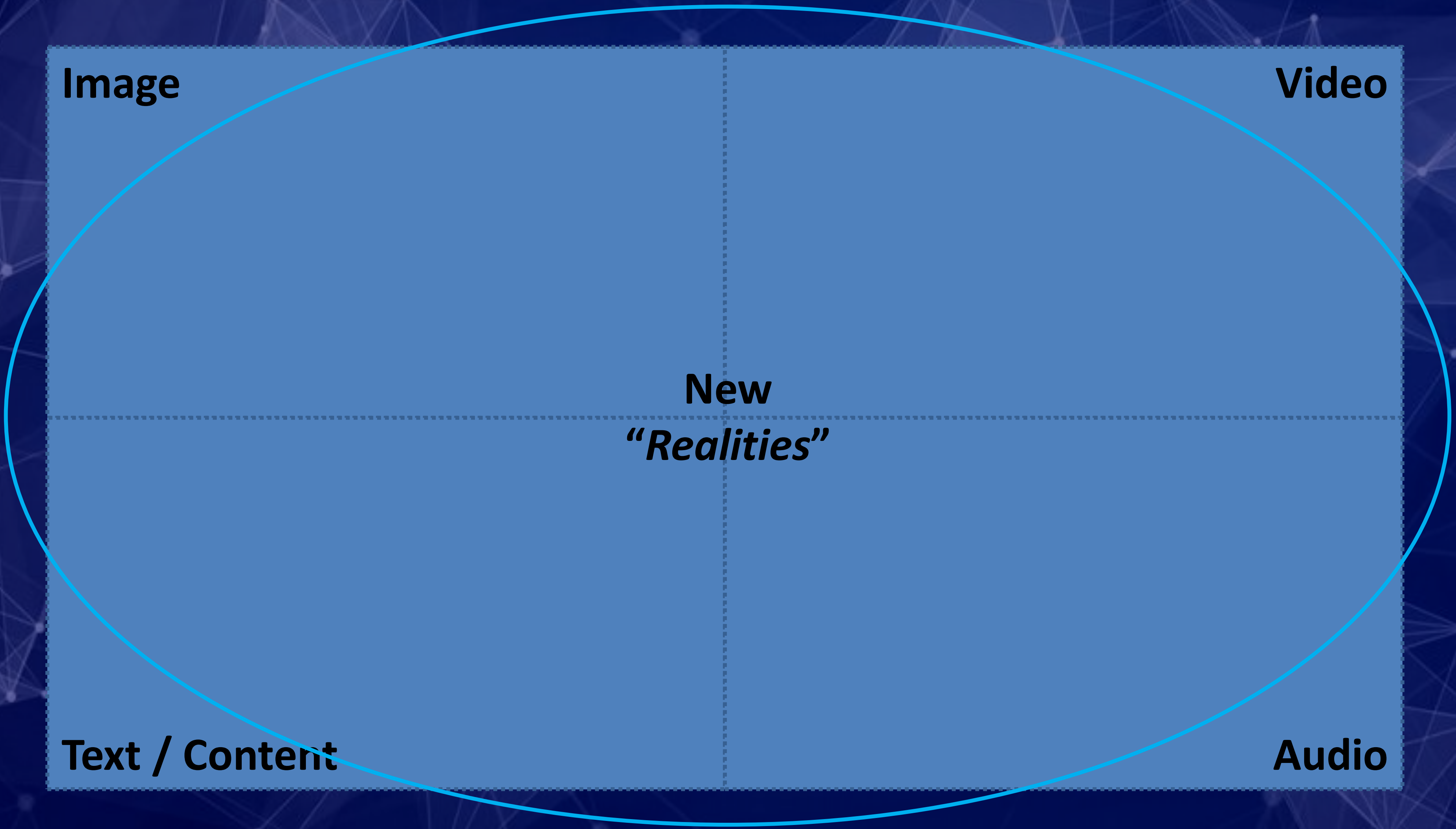
Language Translation

Audio Synthesis

Text / Content

Voice

Fear, Uncertainty, Doubt – *Reality*, *Truth*, *Trust*



Image

Video

New

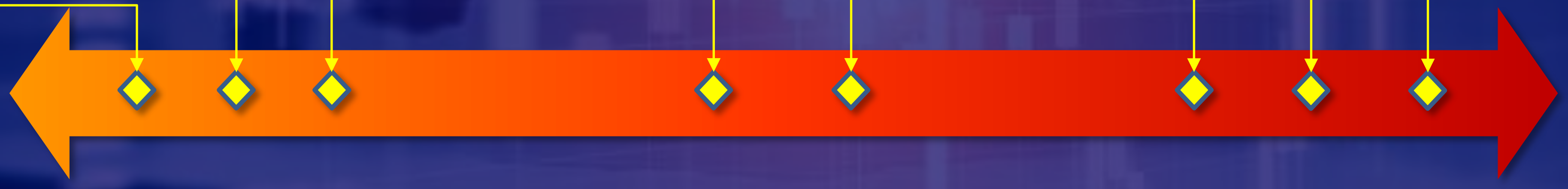
“Realities”

Text / Content

Audio

Phenomenati

Spectrum of Risk Scenarios – for GenAI



Risk #IR001: Confidentiality Loss

Risk #IR002: Poor Integrity of Content

Risk #IR003: Copyright Infringement

Risk #IR004: Availability of the Service

Risk #ER001: Alternative to COMPANY Content or Service

Risk #ER002: More Effective Open Source Recon

Risk #ER003: More Credible Social Engineering attacks

Risk #ER004: More Effective, Polymorphic Malware

Internal Use of External Services

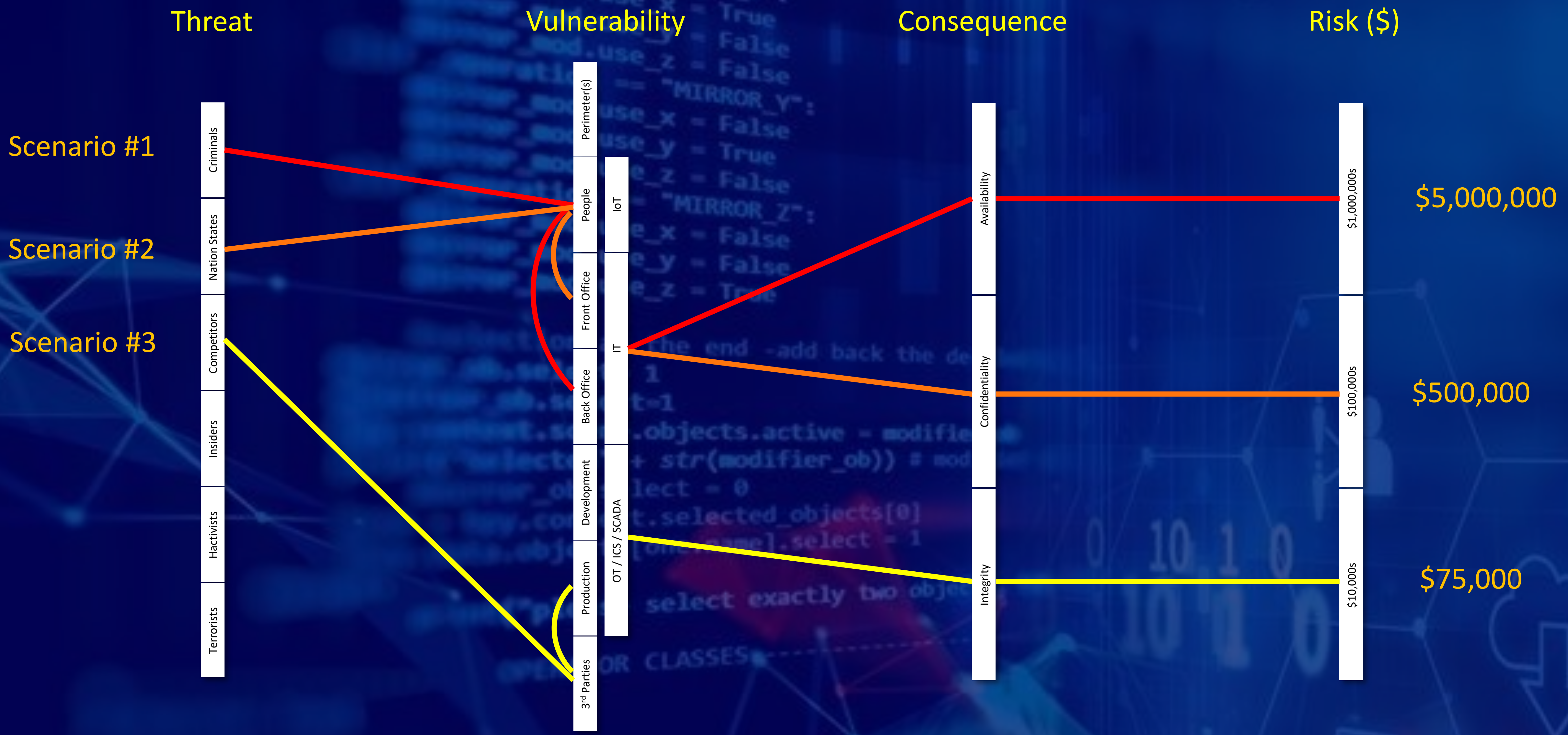
Competitive Threats

External Threats

Leverage GenAI Services in Core Offerings

Risk Scenarios – Connecting the Dots

Phenomenati



Assessment (Qualitative) of Risk Scenarios

	Risk ID	Description	Likelihood	Impact
1	IR001	Loss of **Confidentiality** of Content provided TO Gen AI service(s)	4	3
2	IR002	Poor **Integrity** of Content received FROM Gen AI service(s)	3	5
3	IR003	Content received FROM Gen AI service(s) may violate **Copyrights**	2	3
1	ER001	Gen AI service(s) selected as an **alternative** to COMPANY Service(s)	4	5
4	IR004	New COMPANY Offering/Service becomes critically dependent on **Availability** of Gen AI service	4	4
2	ER002	Threat actors use Gen AI to exploit open source intel for **Reconnaissance** on your staff, business, customers	4	3
3	ER003	**Social Engineering attacks** (phishing, smishing, vishing, live, etc.) are becoming much more effective	3	4
4	ER004	**Malware** is being rapidly **refactored** and **enhanced** (e.g., polymorphic improvements)	5	5

Likelihood	5	5	10	15	20	25
	4	4	8	12	16	20
	3	3	6	9	12	15
	2	2	4	6	8	10
	1	1	2	3	4	5
	4	1	2	3	4	5
		Impact				

Cyber Risk Landscape

Scenario IR001 – Matrix of Potential Controls

Phenomenati

Loss of **Confidentiality** of Content provided TO Gen AI service(s)

IR001 Control Matrix	**Preventative**	**Detective**	**Corrective**
Administrative	**Third Party Risk Management (TPRM)** **Acceptable Use Policy** (of ChatGPT or other Generative AI services) **Training Procedure** on use of Gen AI services	TBD	HR **Disciplinary** Procedure
Technical	**Block** Access to ChatGPT	**Monitor** Access to ChatGPT	TBD
Physical	<not relevant>	<not relevant>	<not relevant>

Scenario IR002 – Matrix of Potential Controls

Phenomenati

Poor **Integrity** of Content received FROM Gen AI service(s)

IR002 Control Matrix	**Preventative**	**Detective**	**Corrective**
Administrative	**Third Party Risk Management (TPRM)** **Acceptable Use Policy** (of ChatGPT or other Generative AI services) **Training Procedure** on use of Gen AI services	New **Review** Processes	Add **Citation** of Source HR **Disciplinary** Procedure
Technical	**Block** Access to ChatGPT	**Monitor** Access to ChatGPT **Scan** products for possible AI generated content	TBD
Physical	<not relevant>	<not relevant>	<not relevant>

Scenario IR003 – Matrix of Potential Controls

Content received FROM Gen AI service(s) may violate **Copyrights**

IR003 Control Matrix	**Preventative**	**Detective**	**Corrective**
Administrative	**Third Party Risk Management (TPRM)** **Acceptable Use Policy** (of ChatGPT or other Generative AI services) **Training Procedure** on use of Gen AI services	New **Review** Processes	Add **Citation** of Source HR **Disciplinary** Procedure
Technical	**Block** Access to ChatGPT	**Monitor** Access to ChatGPT **Scan** products for possible **plagiarism**	TBD
Physical	<not relevant>	<not relevant>	<not relevant>

Scenario ER001 – Matrix of Potential Controls

Phenomenati

Gen AI service(s) selected as **Competition** to COMPANY Service

ER001 Control Matrix	**Preventative**	**Detective**	**Corrective**
Administrative	TBD	Perform **Market Research** to quantify the frequency and demographics of lost business	Adapt **Pricing** strategy **Legal** action to address COMPANY Copyright violations
Technical	Launch **COMPANY Branded** Alternative Service to **ChatGPT**	Aggressive **DRM Program** to detect unauthorized use of COMPANY Content by Competitive Gen AI services	TBD
Physical	<not relevant>	<not relevant>	<not relevant>

Scenario IR004 – Matrix of Potential Controls

New COMPANY Offering/Service becomes critically dependent on **Availability** of Gen AI service

IR004 Control Matrix	Preventative	Detective	Corrective
Administrative	**Third Party Risk Management (TPRM)** **Alternative Providers** of core GenAI Service(s) **Service Level Agreements (SLA)** For Generative AI services	Service **Monitoring & Reporting** Processes	Contractual **Remedies** for missed SLAs **Failback** to **manual** (Real Intelligence) content generation **Insurance** coverage to address Operational outages
Technical	**Deploy** Gen AI service(s) on "internal" infrastructure with **redundant** capacity, high-availability, plan for **Resiliency**, etc.	**Monitor** Availability of the ChatGPT or GenAI service	Automated **Failover** between Gen AI Service(s) or Service Providers Automated **Restoration** of internally hosted Gen AI Service(s)
Physical	<not relevant>	<not relevant>	<not relevant>

Scenario ER002 – Matrix of Potential Controls

Phenomenati

Gen AI service(s) Enables More Effective Open Source Reconnaissance on the COMPANY

ER002 Control Matrix	**Preventative**	**Detective**	**Corrective**
Administrative	Update **Social Media Policy** Update **Security Awareness Training** w/ OPSEC	Regular **Surface Web** scanning Regular **Dark Web** scanning	**Scrub** Social Media **Disciplinary Action** for Policy Violations
Technical	**Poison Open Source** information about the COMPANY (Dis-Information, Deception)	Select/Deploy **Surface Web** scanning services Select/Deploy **Dark Web** scanning services	Select/Deploy **Scrub** Social Media services
Physical	<not relevant>	<not relevant>	<not relevant>

Scenario ER003 – Matrix of Potential Controls

Gen AI service(s) Enables More Credible Social Engineering Attacks on the COMPANY

ER003 Control Matrix	**Preventative**	**Detective**	**Corrective**
Administrative	Update **Security Awareness Alerts & Training** Institute **Authentication** Protocols (e.g., Challenge-Response) Institute **Two-Party Authorization** Protocols (e.g., requires multiple approvals)	Institute **Authentication** Protocols (e.g., Challenge-Response) Establish **Dis-Information, Deception** protocols	Encourage **Aggressive Reporting** behaviors by staff Increase **Threat Sharing** w/ ISACs & Authorities
Technical	Increase sophistication of **email Security Gateways** **MFA** for everything. Accelerate frequency of **credential rotation** Invest in greater **Network Segmentation**	**Monitor for use of Dis-Information, Deception seeds** (e.g., fake links and credentials leading to honeypot, honeynet, sandbox environment) Monitor for **contextually anomalous** behavior	Aggressive **System, Account Isolation**
Physical	<not relevant>	<not relevant>	<not relevant>

Scenario ER004 – Matrix of Potential Controls

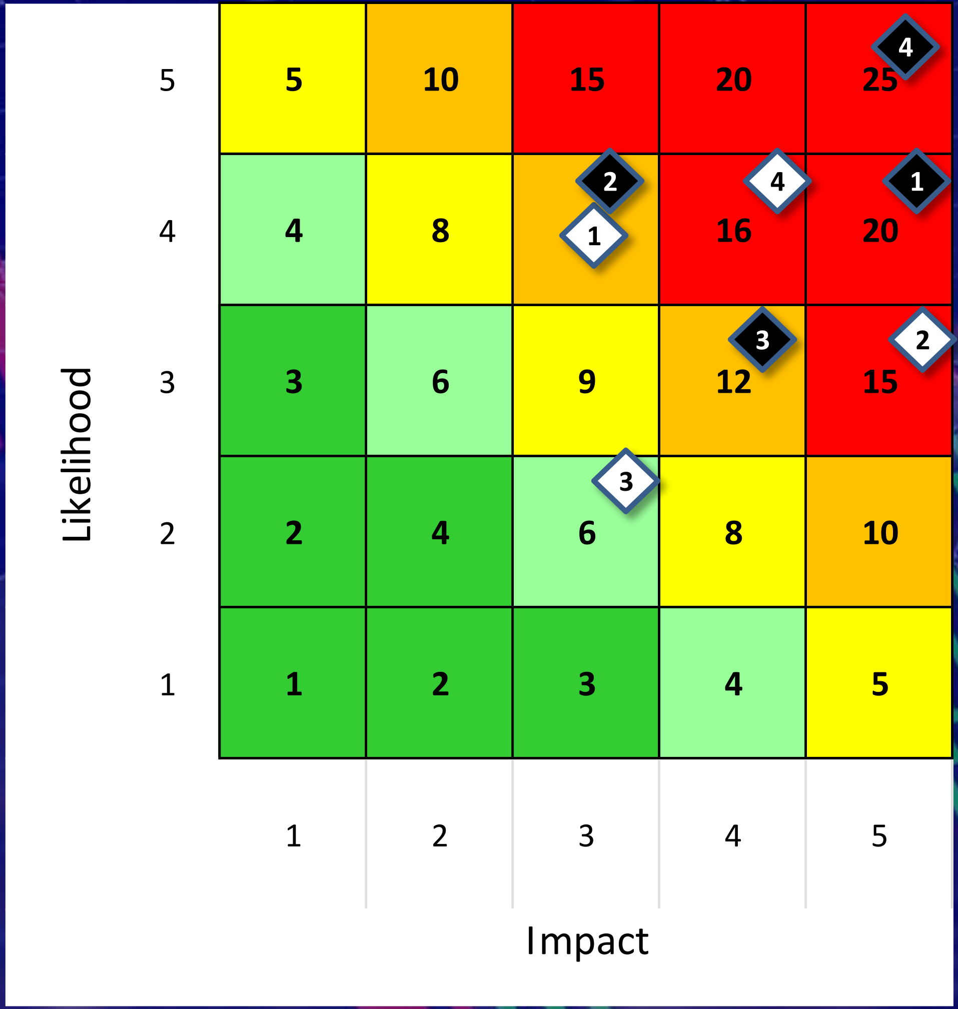
Phenomenati

Gen AI service(s) Enables More Effective Malware Attacks on the COMPANY

ER004 Control Matrix	Preventative	Detective	Corrective
Administrative	Update **Security Awareness Alerts & Training** Increase **Vulnerability** Scanning & **Patch** Management discipline	Fortify the Security Operations Center (SOC) capabilities	Increase **Threat Sharing** w/ ISACs & Authorities
Technical	Strict Enforcement of **Least Privilege** & **Privileged Access Management** services	Employ **Gen AI services** that collect & curate **Threat Intelligence** Continuous updates to **Threat Intel** feeding **EDR**/MDR/NDR/XDR Invest in **UEBA** services	More Aggressive Automatic Enforcement of **Quarantine** & **Isolation** Policies
Physical	<not relevant>	<not relevant>	<not relevant>

Assessment of Risk Scenarios – *Before* Controls

Risk ID	Description	Likelihood	Impact
◇1 IR001	Loss of **Confidentiality** of Content provided TO Gen AI service(s)	4	3
◇2 IR002	Poor **Integrity** of Content received FROM Gen AI service(s)	3	5
◇3 IR003	Content received FROM Gen AI service(s) may violate **Copyrights**	2	3
◆1 ER001	Gen AI service(s) selected as an **alternative** to COMPANY Service(s)	4	5
◇4 IR004	New COMPANY Offering/Service becomes critically dependent on **Availability** of Gen AI service	4	4
◆2 ER002	Threat actors use Gen AI to exploit open source intel for **Reconnaissance** on your staff, business, customers	4	3
◆3 ER003	**Social Engineering attacks** (phishing, smishing, vishing, live, etc.) are becoming much more effective	3	4
◆4 ER004	**Malware** is being rapidly **refactored** and **enhanced** (e.g., polymorphic improvements)	5	5



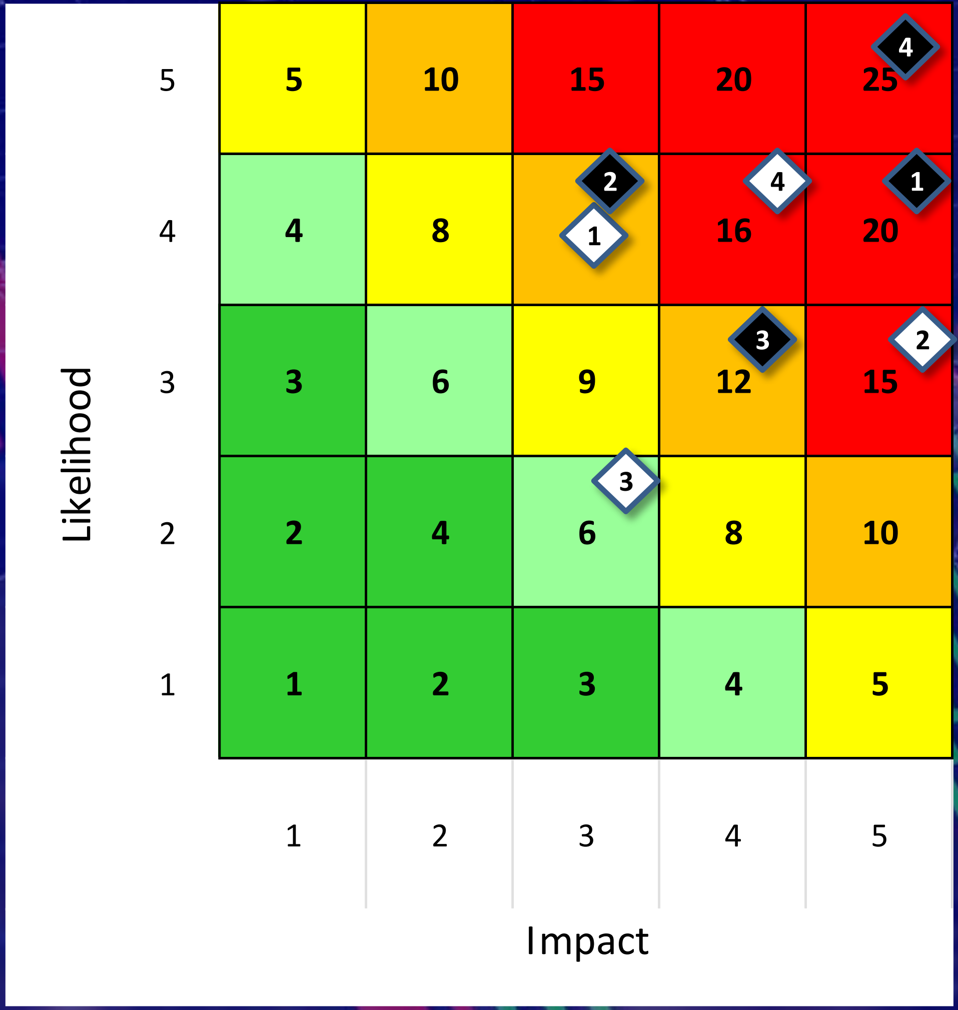
Likelihood					
5	5	10	15	20	25 ◆4
4	4	8	◆2 ◇1	◇4 16	◆1 20
3	3	6	9	◆3 12	◇2 15
2	2	4	◇3 6	8	10
1	1	2	3	4	5
	1	2	3	4	5
	Impact				

Cyber Risk Landscape

Assessment of Risk Scenarios – *After* Controls

Phenomenati

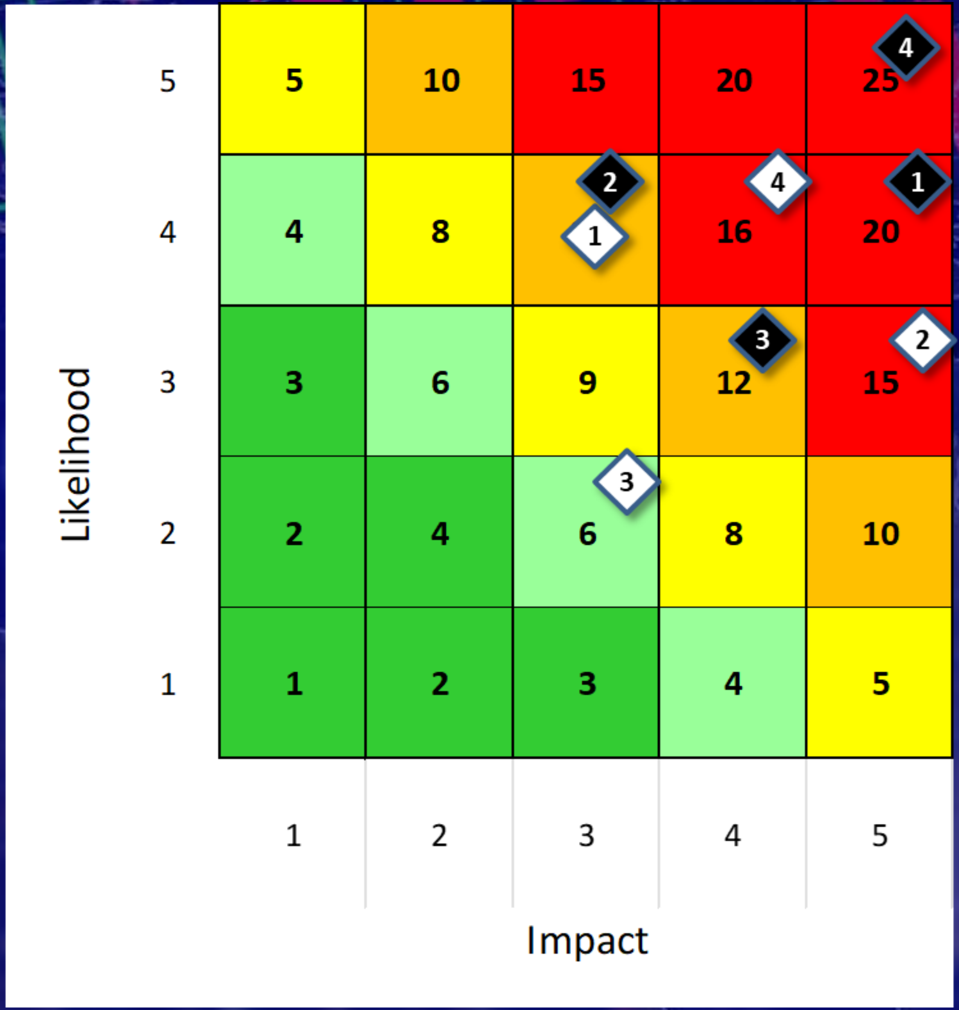
	Risk ID	Description	Likelihood	Impact
1	IR001	Loss of **Confidentiality** of Content provided TO Gen AI service(s)	4 → 1	3
2	IR002	Poor **Integrity** of Content received FROM Gen AI service(s)	3 → 2	5 → 2
3	IR003	Content received FROM Gen AI service(s) may violate **Copyrights**	2	3
1	ER001	Gen AI service(s) selected as an **alternative** to COMPANY Service(s)	4 → 3	5 → 4
4	IR004	New COMPANY Offering/Service becomes critically dependent on **Availability** of Gen AI service	4 → 2	4 → 3
2	ER002	Threat actors use Gen AI to exploit open source intel for **Reconnaissance** on your staff, business, customers	4 → 3	3
3	ER003	**Social Engineering attacks** (phishing, smishing, vishing, live, etc.) are becoming much more effective	3 → 2	4 → 2
4	ER004	**Malware** is being rapidly **refactored** and **enhanced** (e.g., polymorphic improvements)	5 → 4	5



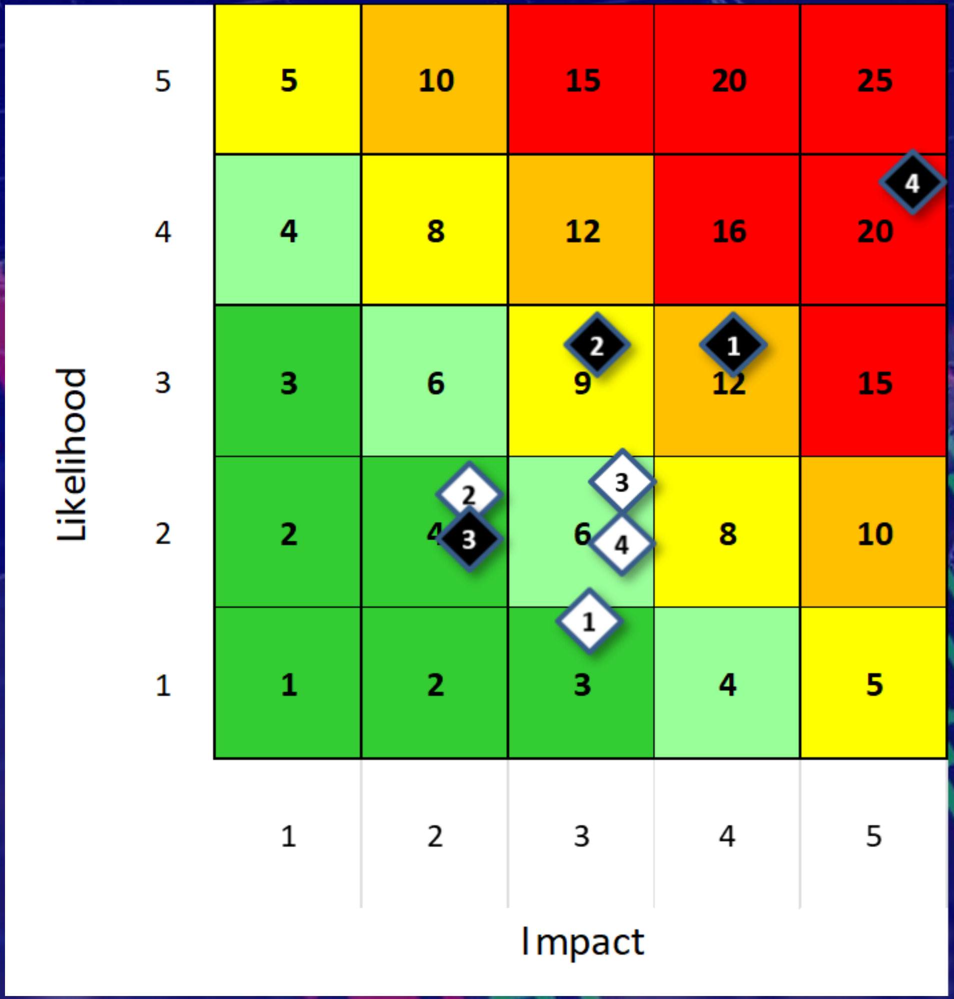
Likelihood					
5	5	10	15	20	25
4	4	8	12	16	20
3	3	6	9	12	15
2	2	4	6	8	10
1	1	2	3	4	5
	1	2	3	4	5
			Impact		

Cyber Risk Landscape

Cyber Risk Landscape – *Before* and *After* New Controls



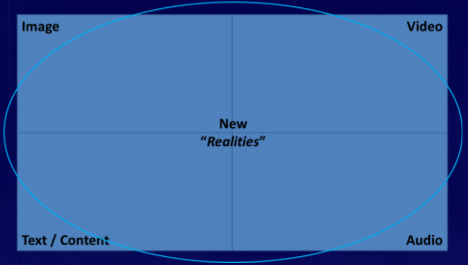
Likelihood					
5	5	10	15	20	25
4	4	8	12	16	20
3	3	6	9	12	15
2	2	4	6	8	10
1	1	2	3	4	5
	1	2	3	4	5
			Impact		



Likelihood					
5	5	10	15	20	25
4	4	8	12	16	20
3	3	6	9	12	15
2	2	4	6	8	10
1	1	2	3	4	5
	1	2	3	4	5
			Impact		

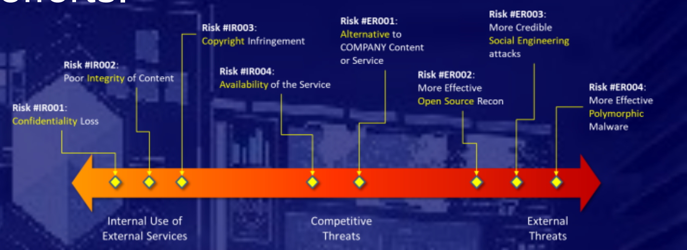
Summary

Summary – Gen AI is *not* a “*Game Changer*”



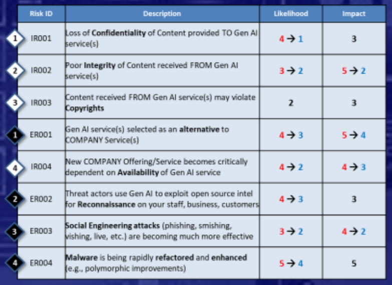
Acknowledge the Patterns… Extinguish the FUD

- AI is not new.
- What IS new is the recent *access* to and *adoption* of Generative AI by everyday users.
- Rapid (“*rabid*”) adoption of new 3rd party services is not new. Update your TPRM efforts.
- Technology as an accelerator and/or market disruptor is not new. Adapt.
- Competition is not new.
- Any Technology has both Legitimate and Malicious applications.

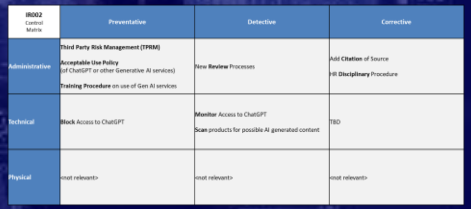

Risk #IR001:
Confidentiality Loss
Risk #IR002:
Poor Integrity of Content
Risk #IR003:
Copyright Infringement
Risk #IR004:
Availability of the Service
Risk #ER001:
Alternative to COMPANY Content or Service
Risk #ER002:
More Effective Open Source Recon
Risk #ER003:
More Credible Social Engineering attacks
Risk #ER004:
More Effective Polymorphic Malware
Internal Use of External Services
Competitive Threats
External Threats

Information is Power

- Dis / Mis-Information has always been a Threat.
- Humans are Vulnerable to Trusting too quickly.
- Gen AI only exacerbates and exploits this.



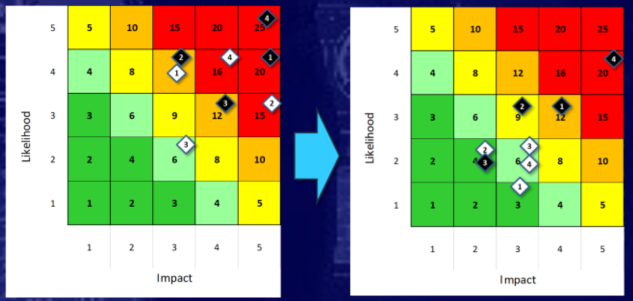
Risk ID	Description	Likelihood	Impact
◇1 IR001	Loss of **Confidentiality** of Content provided TO Gen AI service(s)	4 → 1	3
◇2 IR002	Poor **Integrity** of Content received FROM Gen AI service(s)	3 → 2	5 → 2
◇3 IR003	Content received FROM Gen AI service(s) may violate **Copyrights**	2	3
◆1 ER001	Gen AI service(s) selected as an **alternative** to COMPANY Service(s)	4 → 3	5 → 4
◇4 IR004	New COMPANY Offering/Service becomes critically dependent on **Availability** of Gen AI service	4 → 2	4 → 3
◆2 ER002	Threat actors use Gen AI to exploit open source intel for **Reconnaissance** on your staff, business, customers	4 → 3	3
◆3 ER003	**Social Engineering attacks** (phishing, smishing, vishing, live, etc.) are becoming much more effective	3 → 2	4 → 2
◆4 ER004	**Malware** is being rapidly **refactored** and **enhanced** (e.g., polymorphic improvements)	5 → 4	5



IR002 Control Matrix	Preventative	Detective	Corrective
Administrative	**Third Party Risk Management (TPRM)** **Acceptable Use Policy** (of ChatGPT or other Generative AI services) **Training Procedure** on use of Gen AI services	New **Review** Processes	Add **Citation** of Source HR **Disciplinary** Procedure
Technical	**Block** Access to ChatGPT	**Monitor** Access to ChatGPT **Scan** products for possible AI generated content	TBD
Physical	<not relevant>	<not relevant>	<not relevant>

Bring Order to the Chaos

- Develop Risk Scenarios → Internal, External, Competitive
- Assess the Risk Scenarios
- Develop Recommended Control Matrices to Inform Decision Making
- Communicate using a simple Risk Landscape


Likelihood
Impact
Likelihood
Impact