Generative AI's Impact on Security & GRC

Charles Gillman
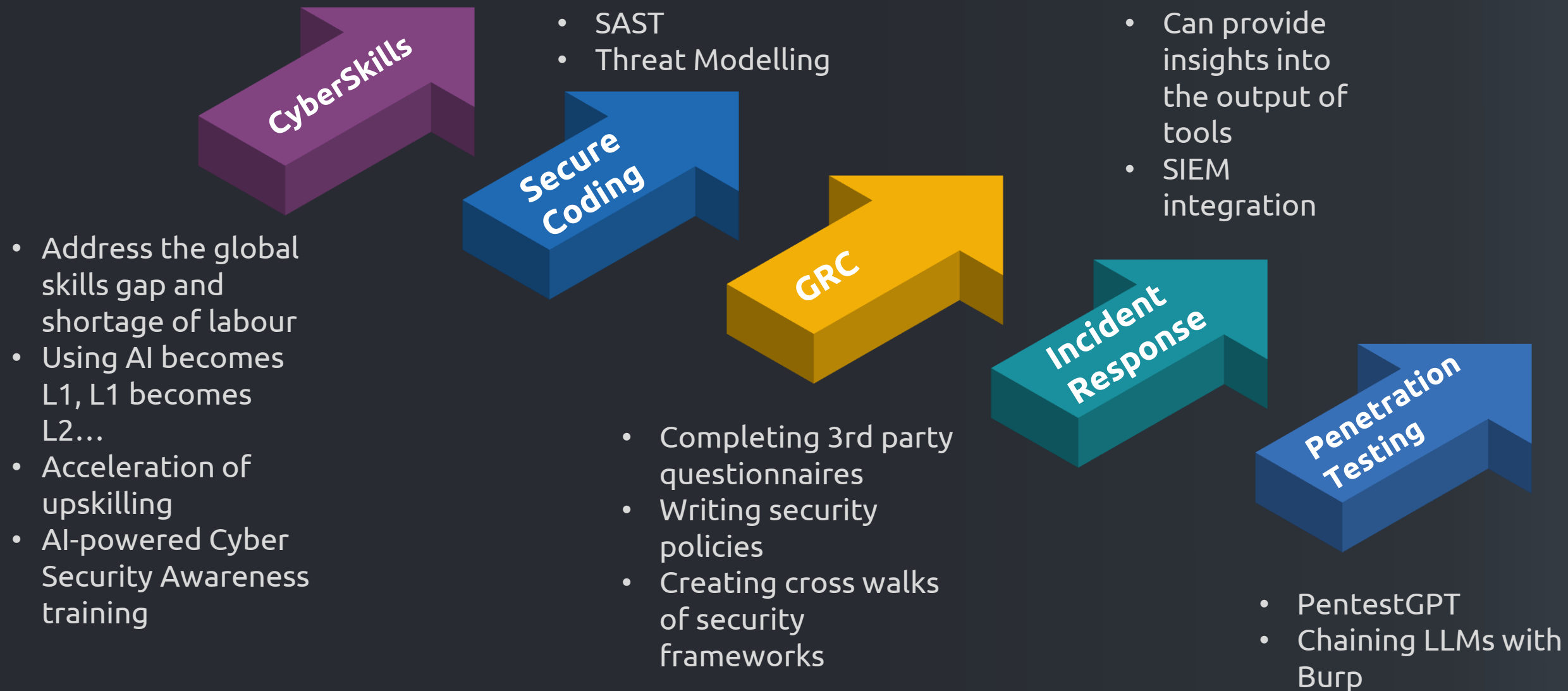
https://www.linkedin.com/in/charles-gillman/

@fullstackciso

The New Frontier in Cybercrime: The Dark Side of AI

Malware Generation

- The attackers started using this to create malware in December 2022
- ChatGPT has already been used to create polymorphic malware

Social Engineering

- Bypass of voice biometrics
- Generating fake Youtube
- Fake news/disinformation
- Impersonation

Exploit Development

- Vulnerability discovery in code
- Creating Exploit code

Phishing/BEC

- Phishing will be grammatically correct. It will be hard to spot
- Conversely, we will be using AI to write emails, making it harder to spot the differences

Cybersecurity Skills Shortage: Can AI Bridge the Gap?

CyberSkills

Secure Coding

GRC

Incident Response

Penetration Testing

- SAST
- Threat Modelling

- Can provide insights into the output of tools
- SIEM integration

- Address the global skills gap and shortage of labour
- Using AI becomes L1, L1 becomes L2...
- Acceleration of upskilling
- AI-powered Cyber Security Awareness training

- Completing 3rd party questionnaires
- Writing security policies
- Creating cross walks of security frameworks

- PentestGPT
- Chaining LLMs with Burp

© Charles Gillman

AI Integration:
Vendors Fast Tracking AI in Cybersecurity Tools

zscaler™
Security Autopilot

CROWDSTRIKE
Charlotte AI (Security Analyst)

Recorded Future®
Analyst GPT

Microsoft
Security Copilot

SentinelOne™
Threat Hunting

ZEROFOX®
FoxGPT Intel Analysis

Google Cloud
Security Analyst Workbench

VERACODE
Fix AI

TITANIAM
Governance for AI

Embracing AI: Balancing Security & GRC

BYOAI

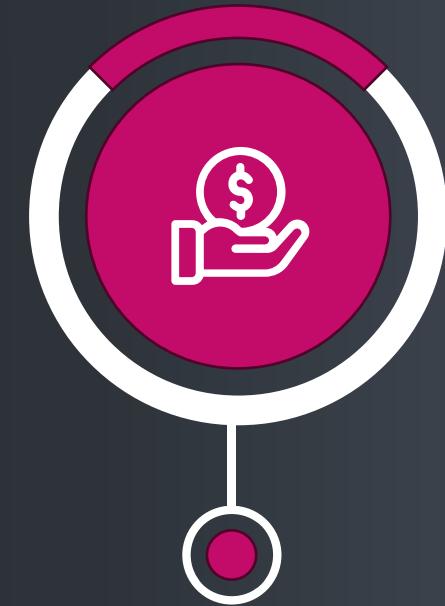AI is becoming ubiquitous

Implementer

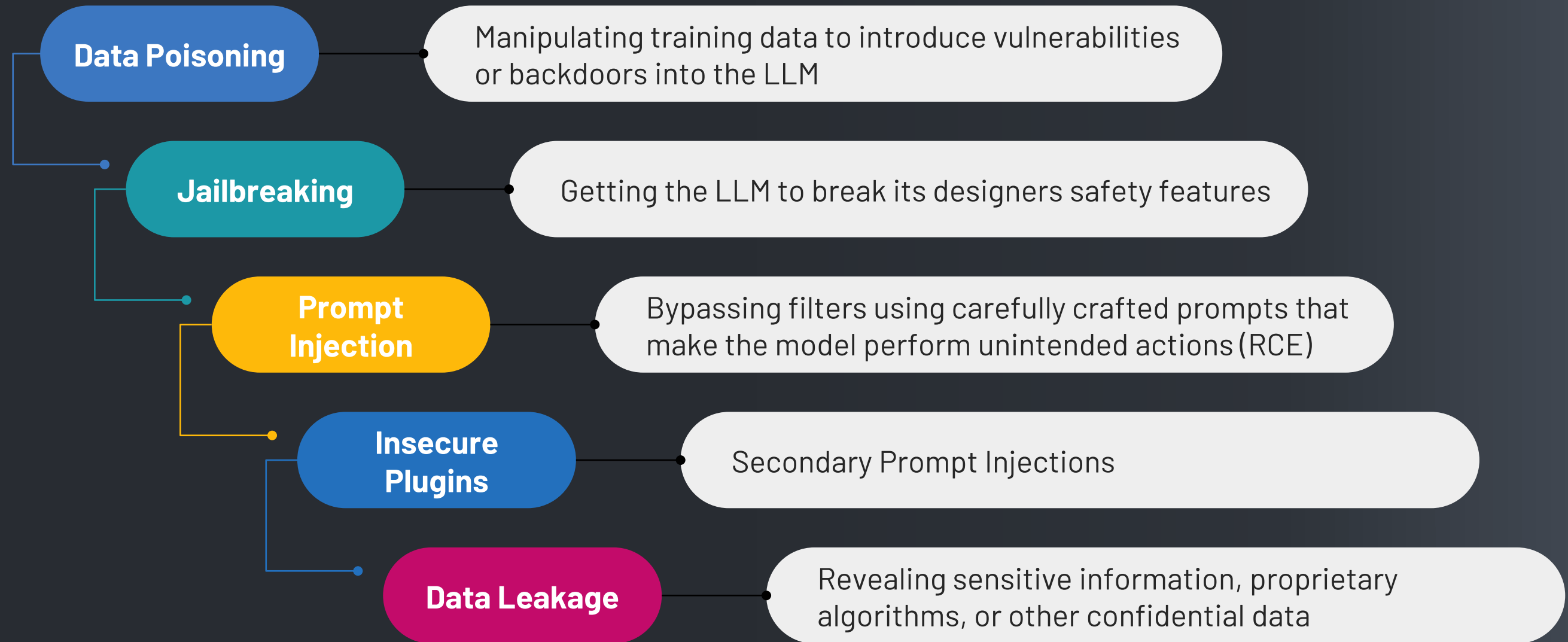Security, Privacy and GRC implications for customer data

Consumer

Security, Privacy and GRC implications for corporate data
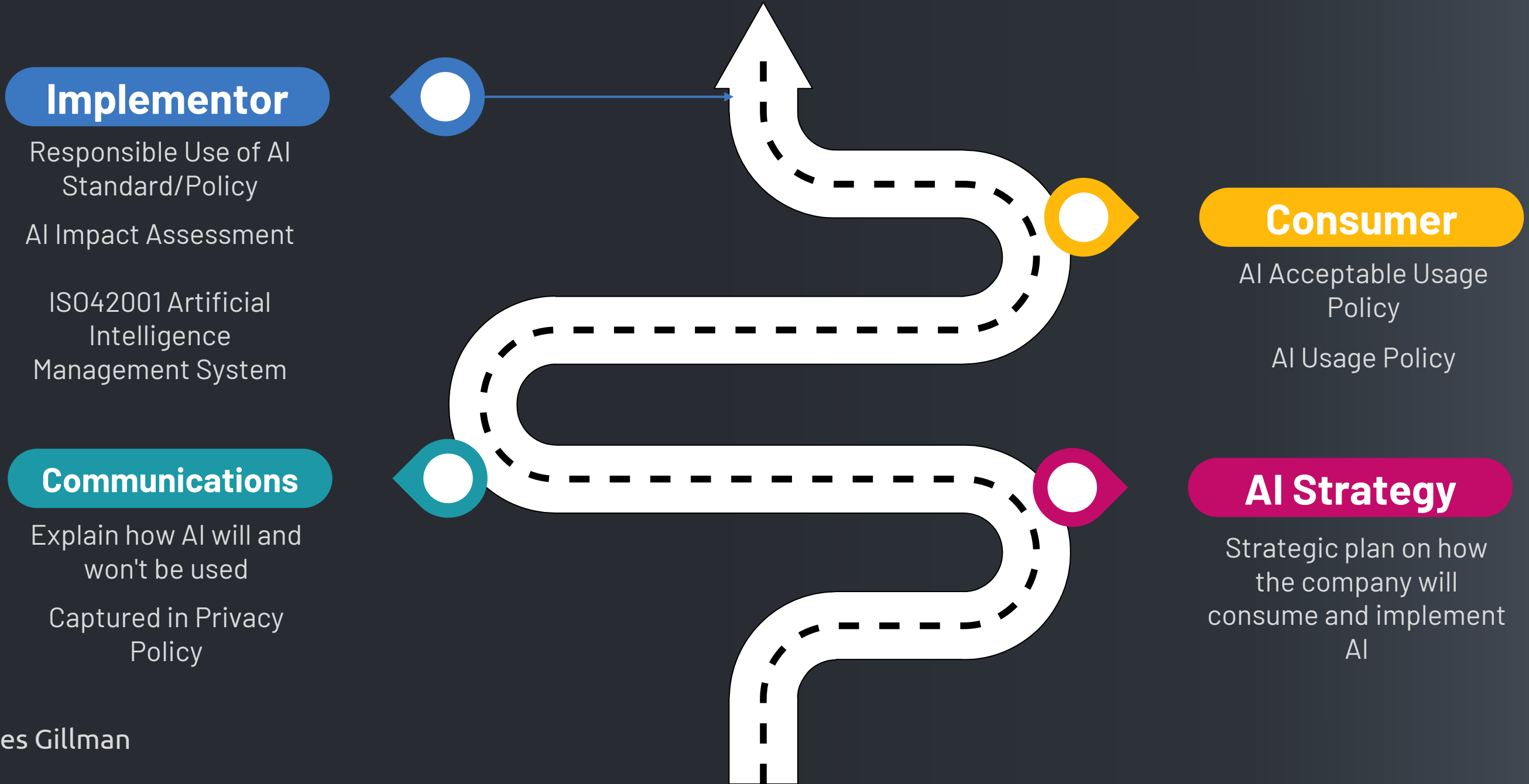
Vendor AI Risk

3rd Party AI Risk Assessment

Trust: Generative AI's Achilles' Heel

Data Poisoning — Manipulating training data to introduce vulnerabilities or backdoors into the LLM

Jailbreaking — Getting the LLM to break its designers safety features

Prompt Injection — Bypassing filters using carefully crafted prompts that make the model perform unintended actions (RCE)

Insecure Plugins — Secondary Prompt Injections

Data Leakage — Revealing sensitive information, proprietary algorithms, or other confidential data

Risk Management: In the AI Era

Consumer Risks

- Bias
- Hallucinations
- Decisions based on output

Implementor Risks

- The AI Attack Surface Map
- MITRE Atlas
- OWASP Guide to AI Security & Privacy
- OWASP LLM Top 10
- Cloud Security Alliance Security Implications of using ChatGPT
- Team 8 Generative AI Enterprise Risks

Manage the Risks

- MITRE Sensible Regulatory Framework for AI
- NIST AI Risk Management Framework
- Private AI Models

Recap:

Weaponising AI — **Cybercriminals**

Cyber Vendors — Embedding AI

Leverage AI to upskill — **Cyberskills**

AI

Risk Manage — Use defined AI RM Frameworks

Define AI Policies — **Govern**

Comply — Track evolving Regulations

© Charles Gillman